THE BIOCHEMICAL INVESTIGATION AND ISOLATION OF SMALL MOLECULE INHIBITORS FOR TWO ESSENTIAL PROTEINS OF *Mycobacterium tuberculosis* H37Rv: IspD AND Wag31

A Thesis

by

SONIA JOSEPH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	James C. Sacchettini
Committee Members,	Gary Kunkel
	Michael Polymenis
	Daniel Romo
Head of Department,	Gregory Reinhart

August 2014

Major Subject: Biochemistry

ABSTRACT

Tuberculosis is one of the leading causes of death due to infectious disease. The causative agent, *Mycobacterium tuberculosis*, is a facultative intracellular parasite with a slow regeneration rate. Though there is a decline in the overall TB incidence since 2005, the emergence of resistant strains that are impervious to existing treatment regimens make the discovery and development of new drug leads crucial. To this end, exploiting key differences between the biology of the host and the pathogen can generate novel lead molecules with minimal side effects. This thesis details the study of two proteins that are essential for the survival of *M. tuberculosis* (M.tb) but are not present in the host, making them potential drug targets.

The first protein, IspD (2-C-methyl-D-erythritol-4-phosphate (MEP) cytidyl transferase), is a part of the non-mevalonate pathway for the synthesis of isoprenoids and catalyzes the condensation of MEP and cytidine triphosphate (CTP) to form 4-diphosphocytidyl-2-C-methylerythritol (CDP-ME) and pyrophosphate (PPi). A medium-throughput enzyme assay was developed to identify inhibitors for this protein. It was screened against 3550 compounds drawn from five different M. tb whole-cell active small molecule libraries generating a total of five hits. These molecules were then assessed for their potency against IspD as measured by their $IC_{50,}$ their activity against M. tb whole cells and their cytotoxicity. Of the five hits, two compounds inhibited M. tb whole cell growth at a concentration below 50 μM while exhibiting no general cytotoxicity to human dermal

fibroblasts (HDF). They each had an IC_{50} of 26.1 µM and 37.8µM and preliminary SAR studies were performed on the latter. These molecules could prove to be a viable starting point for the rational design of IspD inhibitors.

The second protein, Wag31 is a cell division associated protein that regulates mycobacterial cell size and septum formation. Wag31 exhibited a propensity for gel formation both alone and in association with other cellular proteins. A purification strategy was developed to circumvent this tendency and generate soluble protein. It was found that mutations within the wag31 protein coding sequence conferred resistance to a whole-cell active small molecule (MIC_{99}=6.25 µg/ml) in both *M. tuberculosis* and *M. smegmatis*. Moreover, all of the discovered mutations were clustered within the C-terminal coiled-coil domain of the protein. It was established that this compound binds to Wag31 and seems to shift the equilibrium of the protein solution towards gel formation. The mutated protein does not form gel and seems to bind to the compound at a significantly reduced rate. To further confirm that Wag31 was indeed the target of this small molecule, whole cell viability assays were performed to establish whether the over-expression of Wag31 in *M. smegmatis* would shift the EC_{50} and MIC_{99} values. Wag31 over-expression reduced both the EC_{50} and MIC_{99} values providing further proof that Wag31 is the target for this compound. The compound appears to act by shifting the equilibrium of the protein towards a gelatinous state which proves inhibitory for cell growth.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sacchettini, and my committee members, Dr. Kunkel, Dr. Polymenis and Dr. Romo for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues, in particular Jennifer Tsai and Mallikarjun Lalgondar, and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my family for their encouragement and to my husband for his patience and love.

NOMENCLATURE

MESG	2-amino-6-mercapto-7-methylpurine riboside
MEP	2-C-methyl-D-erythritol-4-phosphate
IspD	2-C-methyl-D-erythritol-4-phosphate (MEP) cytidyl transferase
CDP-ME	4-diphosphocytidyl-2-C-methylerythritol
A	Alanine
R	Arginine
N	Asparagine
D	Aspartate
C	Cysteine
CTP	Cytidine Triphosphate
DSF	Differential scanning fluorimetry
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
E	Glutamate
Q	Glutamine
G	Glycine
$MC^2$7000	M.tb H37Rv BCG strain
$MC^2$4517	*M. smegmatis* competent cell strain
EC_{50}	Half maximal effective concentration of a drug
H	Histidine

IC_{50}	Half maximal inhibitory concentration of a drug
I	Isoleucine
IPTG	Isopropyl-β-D-1-thio-galactopyranoside
L	Leucine
K	Lysine
M	Methionine
MIC	Minimum inhibitory concentration of a drug
M. tb	*Mycobacterium tuberculosis*
F	Phenylalanine
PCR	Polymerase chain reaction
P	Proline
PPi	Pyrophosphate
RBS	Ribosome silencing factor
S	Serine
T	Threonine
W	Tryptophan
TB	Tuberculosis
Y	Tyrosine
V	Valine

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iv

NOMENCLATURE v

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES xi

CHAPTER I INTRODUCTION 1

1.1 Tuberculosis: the disease and its challenges 1
1.2 IspD (2-c-methyl-d-erythritol-4-phosphate (MEP) cytidyl transferase) 2
1.3 Wag31 8

CHAPTER II SMALL MOLECULE SCREENING OF IspD: RESULTS 13

2.1 The development of an assay to measure IspD enzyme activity 13
2.2 The adaptation of the MESG assay for IspD to a 384-well format 16
2.3 The screening results 19
2.4 Determination of IC_{50} 21
2.5 Whole cell assays 23
2.6 Exploring chemical space with structural analogs 25
2.7 Crystallization trials 26

CHAPTER III PURIFICATION AND CHARACTERIZATION OF Wag31: RESULTS 29

3.1 Purification 29
3.2 Resistant mutant studies 32
3.3 Binding studies 34
3.4 Whole cell viability assays 36
3.5 Gel formation in the presence and absence of compound 38

CHAPTER IV MATERIALS AND METHODS....43

4.1 Cloning, protein over-expression and purification....43
4.2 Protein crystallization trials....46
4.3 Enzyme assays, screening and dose response curves....47
4.4 Whole cell assays....47
4.5 Intrinsic fluorescence measurement....48
4.6 Mutation studies....49
4.7 Gel harvesting and pull-down assays....49
4.8 Western blot and protein quantification....50

CHAPTER V SUMMARY AND CONCLUSION....51

REFERENCES....53

LIST OF FIGURES

Page

Figure 1. Schematic representation of the mevalonate and non-mevalonate pathways.....4

Figure 2. Schematic representation of the reaction catalyzed by IspD 7

Figure 3. MESG assay for the detection of pyro phosphate.....14

Figure 4. Optimization of IspD enzyme assay 15

Figure 5. IspD enzyme activity monitored over a 15 minute interval.....16

Figure 6. Scatter plot representing IspD screened against SAC1 diversity library plate no.6085…..…… 17

Figure 7. Scatter plot representing IspD screened against SRI library plate 2.....18

Figure 8. Final hits from the IspD enzyme screen 20

Figure 9. Dose response curve for (a) SRIa (b) SAC1c (c) SAC1a (d) SAC1b.....21

Figure 10. $MC^2$7000 growth inhibition by IspD screen hits 24

Figure 11. Human dermal fibroblasts growth inhibition by IspD screen hits 25

Figure 12. Hits obtained using structure similarity search on Scifinder 26

Figure 13. IspD crystals and diffraction patterns. 28

Figure 14. SDS gel seperation of lysate and supernatant fractions of Wag 31 30

Figure 15. SDS gel separation of Ni affinity chromatography wash and gradient fractions obtained on purifying Wag31. 30

Figure 16. Western blot analysis of Wag31 Ni chromatography fractions 31

Figure 17. Compounds that led to the isolation of $MC^2$7000 resistant mutants with point mutations in the Wag31 protein coding region 32

Figure 18. Intrinsic fluorescence measurement of Wag31.....35

Figure 19. Concentration of Wag31 with 10 M excess of compound A......35

Figure 20. Whole cell growth inhibition assay in $MC^2$4517, $MC^2$4517 RSF OE and $MC^2$4517 Wag31 OE 36

Figure 21. Relative amounts of gel in the supernatant after lysing the cells overexpressing the three Wag31 variants 39

Figure 22. Protein concentrated in the presence of compound A and the respective flow through.....40

Figure 23. Protein gel formed in the absence and presence of compound A.....41

Figure 24. SDS gel separation of pull-down performed using the Wag31 gel as bait.....42

LIST OF TABLES

Page

Table 1. Summary of Wag31 constructs and the behavior of the corresponding proteins on purification. 31

Table 2. List of $MC^2$7000 compound A resistant mutants with the corresponding mutations identified 33

Table 3. Summary of EC_{50}, MIC_{90} and MIC_{99} values obtained from whole cell growth inhibition assays 38

Table 4. List of PCR primers used to generate Wag31 constructs 43

Table 5. List of PCR primers used to generate Wag31 mutants 44

CHAPTER I

INTRODUCTION

1.1 Tuberculosis: the disease and its challenges

Tuberculosis (TB) is an airborne infection caused by various species of mycobacteria, primarily *Mycobacterium tuberculosis*, which claimed up to 1.4 million lives in 2010. It is a disease that affects the developing world with a disproportionate number of its victims being young people in their prime. Although there has been a steady decline in the TB incidence rate and death rate since 2005, the rise in the reported cases of drug-resistant strains pose new challenges in the elimination of this pathogen [1]. The emergence of resistant strains of M. tb is both a function of the extensive duration of chemotherapy as well as the persistent nature of the infection. Some infections are known to spontaneously reactivate after 6 months of medical intervention [2]. The current treatment involves the administration of Isoniazid (INH), Rifampicin (RIF), Pyrazinamide (PZA) and Ethambutol (EMB) in combination for 2 months followed by 4 months of INH and RIF [3]. Multi-drug resistant (MDR) TB strains exhibit resistance to at least INH and RIF which necessitates treatment with second line drugs like fluoroquinolones, Ethionamide (ETH), Kanamycin, Capreomycin and D-cycloserine [2]. This can not only increase the cost of treatment (200 fold) and its side effects; MDR TB chemotherapy can last up to 2 years [4]. According to the WHO Global TB Control Report of 2011, MDR TB accounted for 650,000 of the TB deaths in 2010 [1]. Resistance to fluoroquinolones and one or more of the other second line drugs in

addition to the core TB drugs has been reported in at least 77 countries and is categorized as XDR (eXtremely Drug Resistant) TB by the WHO [5]. There have been three reported outbreaks of TDR (Totally Drug Resistant) TB so far: in Italy in 2007 [6], in Iran in 2009 [7] and in India in 2011 [8]. The most effective treatment for TB was introduced half a century ago and by 2012, resistance to all the existing drugs has already been documented. It is clear that M. tb is a formidable opponent and understanding its biology and isolating new drug targets to overcome the existing resistant mutations is of the greatest importance.

A 2008 *in silico* high throughput target identification study revealed potential M. tb drug targets by selecting for a range of properties including but not limited to: essentiality, absence of a homolog in the host and absence of homologs in the proteome of the gut flora. The proteins described in this proposal are two of the enzymes identified by this method making them valuable drug targets [9].

1.2 IspD (2-c-methyl-d-erythritol-4-phosphate (MEP) cytidyl transferase)

The non-mevalonate pathway for the synthesis of isoprenoids

Isoprenoids are a large and diverse group of organic compounds that are derived from five carbon isoprene units. These naturally occurring lipids are of great physiological significance; they are involved in electron transport, photosynthesis, regulating gene expression and signal transduction. They are essential for the formation of membranes and are components of various hormones, vitamins, scents and pigments [10]. In

Mycobacteria, they are involved in the development of the cell wall, including the synthesis of mycolic acids and lipoarabinomannan [11]. Isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are the precursors to all biological isoprenoids. The high profile of isoprenoids in biological processes makes the synthesis of these precursors an essential metabolic event. There are primarily two cellular pathways that are involved: the Mevalonate pathway and the Non-mevalonate/1-deoxy-d-xylulose-5-phosphate (DOXP) pathway/2*C*-methyl-d-erythritol-4-phosphate (MEP) pathway with the latter being a relatively recent discovery [12, 13]. For the longest time, it was assumed that the mevalonate pathway was the only route toward isoprenoid precursors. It is now known that plant chloroplasts, algae, cyanobacteria, prokaryotes and apicomplexan parasites utilize the non-mevalonate pathway [14].

The intermediates generated decide the difference between the two pathways and their names: the mevalonate pathway proceeds via the synthesis of mevalonate while the key intermediate in the non-mevalonate pathway is 2*C*-methyl-d-erythritol-4-phosphate (MEP). (Figure 1)

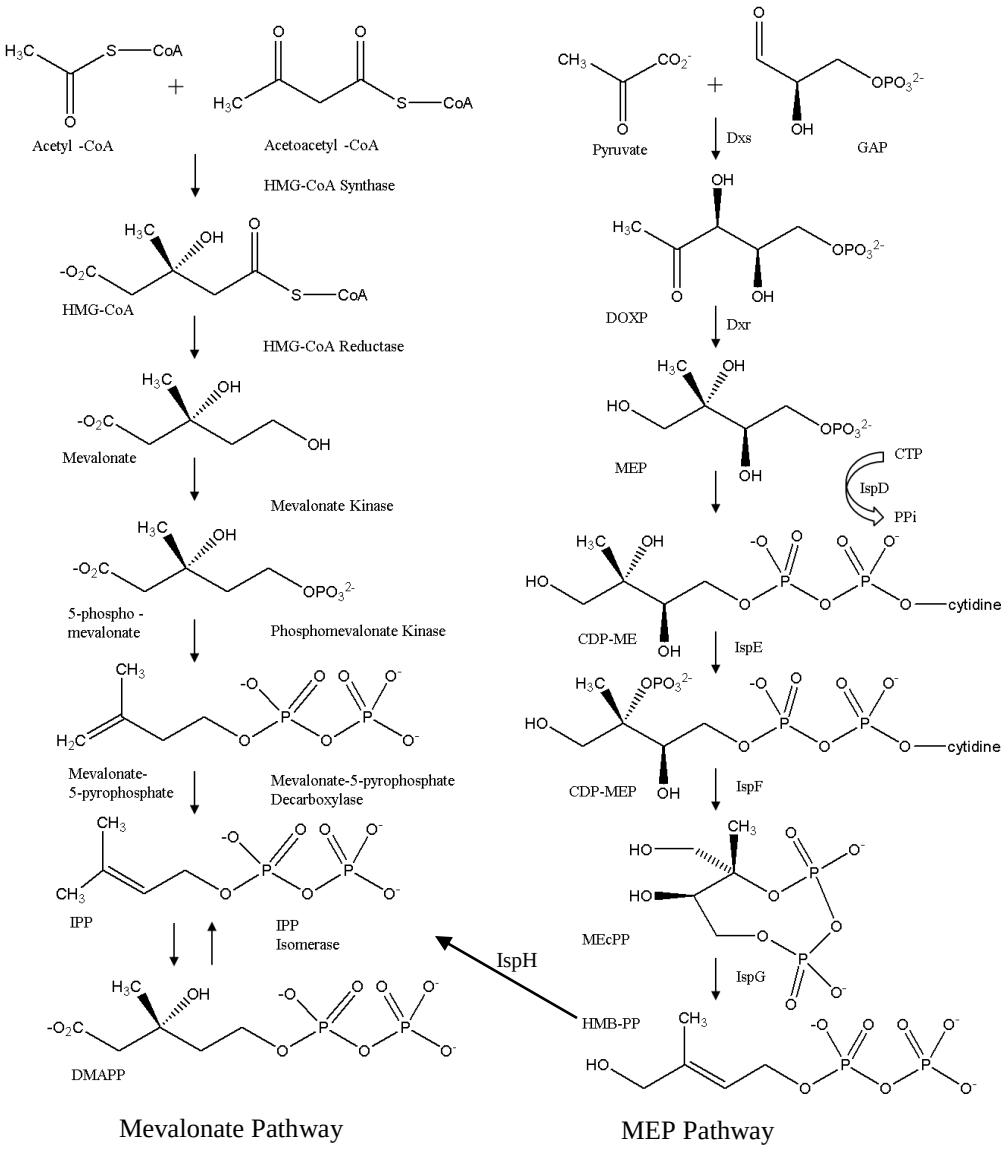


Figure 1. Schematic representation of the mevalonate and non-mevalonate pathways

The MEP pathway proceeds via seven enzymatic reactions that convert pyruvate and glyceraldehyde-3-phosphate to IPP and DMAPP (Figure 1). The first step, catalyzed by the enzyme DXS (1-deoxy-D-xylulose 5-phosphate synthase), condenses pyruvate and glyceraldehyde-3-phosphate to form DOXP (1-deoxy-D-xylulose 5-phosphate).

Subsequently DOXP is converted into MEP (2-C-methyl-D-erythritol-4-phosphate) by DXR (1-deoxy-D-xylulose 5-phosphate reductoisomerase) *via* the oxidation of NADPH. IspD transfers the cytidyl moiety from CTP to MEP to form CDP-ME. The IspE enzyme facilitates the addition of a phosphate to CDP-ME from ATP forming CDP-MEP. CDP-MEP undergoes three more enzymatic steps, (catalyzed by IspF, IspG and IspH) that involve the removal of the nucleotide followed by two consecutive dehydration reactions to form IPP and DMAPP [14].

All of the enzymes in the MEP pathway have been studied extensively. The *E. coli* protein structures have been solved for all the enzymes except IspG [14, 15]. In case of the latter, the *Thermus thermophilus* [16] and *Aquifex aoelicus* [17] homologous protein structures have been solved. *M. tuberculosis* protein structures have been solved for DXR [18], IspD (PDB ID: 30KR) and IspE [19] and the *Mycobacterium smegmatis* IspF [20] structure has been solved.

Since the non-mevalonate pathway is entirely absent in the host (in this case *Homo sapiens*) and the host and pathogen use different pathways to synthesize isoprenoids, the enzymes of this pathway have long been considered attractive targets for drug development [14, 21].

In *M. tuberculosis*, the essentiality of this pathway has been established [22] and provides a hitherto unexplored avenue for drug development. It also furnishes an

alternate route to attack the cell wall which is the target of time tested drugs like INH, ETH and EMB [2].

One of the enzymes in the MEP pathway, DXR, is the well-known target of the *Plasmodium* drug Fosmidomycin, but is not essential for the growth of *M. tuberculosis in vitro* [23]. In addition, Fosmidomycin cannot penetrate the thick cell wall of the pathogen. This makes a strong argument for the development of alternate enzyme targets along this pathway such as IspD.

The IspD enzyme and its value as a drug target

Genetic analysis determined that IspD is encoded by the *M. tuberculosis* gene, Rv3582c, which showed 31% identity to the amino acid sequence of the *E. coli* homolog [24]. It shares the highest sequence similarity to the *M. bovis* gene (99%) with sequence similarity dropping drastically across genera. This gene is believed to be part of a bicistronic operon which also carries the IspF gene (Rv3581c) [22]. The means by which IspD gene expression is regulated are not well characterized.

Rv3582c is 696 bp long and encodes a polypeptide chain comprising 234 amino acids with a molecular weight of 26 kDa. IspD is a cytosolic protein and is known to be a functional dimer [24].

IspD has been demonstrated to be essential for the *in vitro* survival of *M. tuberculosis* both by transposon site hybridization (TraSH) [23] and the inability to isolate a non-functional mutant in H37Rv using gene switching analysis [24]. This further enhances the enzyme's appeal as a drug target.

IspD catalyzes the transfer of the cytidyl moiety from CTP to MEP, thereby forming CDP-ME and releasing PPi and is therefore categorized as a transferase (Figure 2). The reaction is thought to be sequential with CTP binding first followed by MEP. A magnesium ion is required to stabilize the pentacoordinate phosphate transition state that results from the nucleophilic attack of the MEP phosphate on the α-phosphate of the CTP [25].

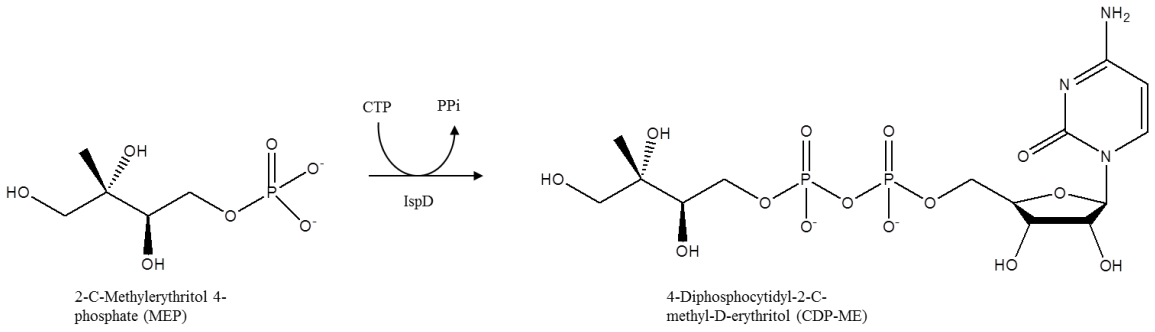


Figure 2. Schematic representation of the reaction catalyzed by IspD

The kinetics of IspD has been the focus of various studies. There is a discrepancy in the values published by the various groups which could be attributed to the different assays and reaction parameters used to measure enzyme activity [24, 26, 27]. However, a repetition of the experiments in our lab re-confirmed a Km of 43.5 for MEP and 45.5 for

CTP as reported in the study by Björkelid *et al* [27]. IspD is active over a broad pH range and the essentiality of a divalent ion was demonstrated by the fact that addition of EDTA completely extinguished its cytidyl transferase activity. It shows better activity with magnesium than with other ions, which is consistent with other IspD homologs [24].

The crystal structure of *M. tuberculosis* IspD has been solved in our lab (courtesy Dr. Manchi Reddy). We have submitted to PDB the structures of the apo-protein at a resolution of 2.4 Å (PDB ID: 30KR), protein bound with the substrate, CTP, at 2.1 Å (PDB ID: 3Q7U) and protein with product, CDP-ME, bound at 2 Å (PDB ID: 3Q80). We have been unable to obtain crystals with MEP bound at the active site which is consistent with the proposed mechanism which requires CTP binding first in order to recruit MEP to the active site [25]. The structure showed significant similarity to the *E. coli* protein, however, there were some significant differences between active site residues that made it more flexible and may help explain the difference in the reaction kinetics between the two homologs.

1.3 Wag31

DivIVA family of proteins

The DivIVA protein is highly conserved in gram positive bacteria. It is a cell division and chromosome partitioning associated protein (NCBI Pubmed Conserved Domain Database). In rod shaped cells, it is known to localize to the cell poles and is believed to

act as a scaffold for other proteins during cell wall biosynthesis and cell division. It is known that DivIVA localization at the poles is due to its recognition of the spherical inward curve of these cells near the poles compared to the remaining portion (which maintains a cylindrical curve which is less sharp). Furthermore, it has been postulated that it is this same negative curvature that guides its localization to septum formation sites during cell division [28]. The loss of DivIVA causes abnormal morphology in these cells, though the nature of this abnormality varies between different species. The proteins of this family also exhibit a pronounced tendency to form multimers [29]. It has been suggested that this oligomerization plays a key role in its localization at the poles of rod-shaped cells [30].

As of April 2014, a single structure has been solved for this family of proteins: that of the *Bacillus subtilis* DivIVA N-terminal domain. Combined with a low resolution structure of the C-terminal domain, a model for the full length protein structure has been proposed [29]. The extremely limited data in this regard ensures that any new structural information regarding members of this family of proteins will be highly valuable.

Wag31: essential and enigmatic

Wag31 is encoded by Rv2145c of the *M. tuberculosis* H37Rv strain. It encodes a polypeptide that is 260 amino acids long with a molecular weight of 28.2 kDa and was found to be essential by TraSh [23]. The *M.smegmatis* ortholog shows 78% sequence identity and 85% similarity over the entire length of the M.tb protein sequence. Attempts

to disrupt the *M. smegmatis* wag31 gene were unsuccessful which supports the importance of this protein for cell survival in Mycobacteria [31].

The link between Wag31 and cell morphology was first illuminated during an investigation of the function of the PknA and PknB threonine kinases which were shown to regulate cell morphology. Wag31 was identified as one of the potential cellular substrates of the PknA kinase and the phosphorylation was found to occur on the threonine 73 residue. It was later determined that phosphorylation state of Wag31 regulates the morphology of the M.tb cell. The current hypothesis maintains that PknB phosphorylates PknA in response to some external stimuli and that the latter protein then proceeds to phosphorylate Wag31. The means by which Wag31 regulates cell morphology is currently not known [32].

Protein structure prediction using MultiCoil algorithm [33] indicates that Wag31 like other DivIVA family proteins has a coiled-coil domain at both C- and N- terminals interspersed by a highly variable region that contains the phosphorylation site [32].

Using *M. smegmatis* as a model, it has also been confirmed that Mycobacterial Wag31 like other DivIVA family proteins localizes at the poles (for the bulk of the cell cycle) and the sites of septum formation (at the time of cell division). What is unusual about this enzyme is that overexpression often leads to accumulation of the enzyme along the cylindrical walls of the cell, stimulating the formation of branches. This seems to

contradict the general assumption that DivIVA family protein localization is based on the recognition of negative cell wall curvature. Overexpression can also lead to the formation of bowling-pin shaped cells with 80 fold the volume of a regular cell and an accumulation of Wag31 at the bulbous end. Cell wall biosynthesis and cell elongation were accelerated in the regions were Wag31 accumulated, while septa formation was inhibited in that vicinity. The irregular septa formation and cell volume interfered with regular cell division and reduced cell viability. The cell is unable to tolerate this phenotype for long and self-regulates Wag31 expression in order to restore the regular cell morphology. Within 5 days of transformation of the overexpression plasmid, a marked decrease in Wag31 expression was observed [31].

Recently, it has been proposed that the differential growth between the phosphorylated and dephosphorylated forms of Wag31 is due to increased peptidoglycan synthesis in the regions where Wag31 is localized. This explains the rod shape of the cells when they are growing normally as well as the branching and bowling pin shaped cells in the overexpression strains. The exact nature of its interaction with peptidoglycan biosynthesis proteins remains to be elucidated [34, 35].

Research indicates that Wag31 expression is positively regulated by the stringent response, suggesting that it may be involved in virulence [36]. Data also suggests that it helps the pathogen withstand oxidative stress [37].

Wag31 is thus a protein that performs a diverse range of important functions in mycobacteria. Combined with its essentiality for mycobacterial survival and the absence of a homolog in the host, it constitutes an ideal target for drug development.

CHAPTER II

SMALL MOLECULE SCREENING OF IspD: RESULTS

2.1 The development of an assay to measure IspD enzyme activity

IspD was purified according to the protocol previously established in the lab. The 2-amino-6-mercapto-7-methylpurine riboside (MESG) assay for the detection of phosphates was used to detect enzyme activity (Figure 3). Protein, substrate and coupled enzyme concentrations were optimized to maximize signal to noise ratio. It was observed that at 60 nM (protein conc.), even with substrate concentrations at or near K_m values, the read-out was high and did not fluctuate sharply (Figure 4a). It was decided to do all further studies using this concentration. IPPase and PNP concentrations were optimized next. IPPase concentration was finalized to 0.1 U/ml and it was determined that PNP concentration at 0.06 U/ml gave the greatest signal to noise separation with the minimal deviation in absorbance values (n=10) (Figure 4b).

This higher separation in absorbance also involved using a near saturating concentration of CTP. Since the reaction mechanism requires the binding of CTP before MEP can bind, it was proposed that a higher occupancy of CTP in the enzyme active sites would be required to propel the reaction forward at a reasonable rate. It has also been advised that in the case of multi-substrate reactions, the mechanism of action may determine the substrate concentrations that are optimal for HTS screening [38]. MESG concentration was optimized to 200 µM. The assay was then monitored over time and it was

determined that it reaches saturation at 15 minutes post substrate addition (Figure 5, all experiments were performed in duplicates).

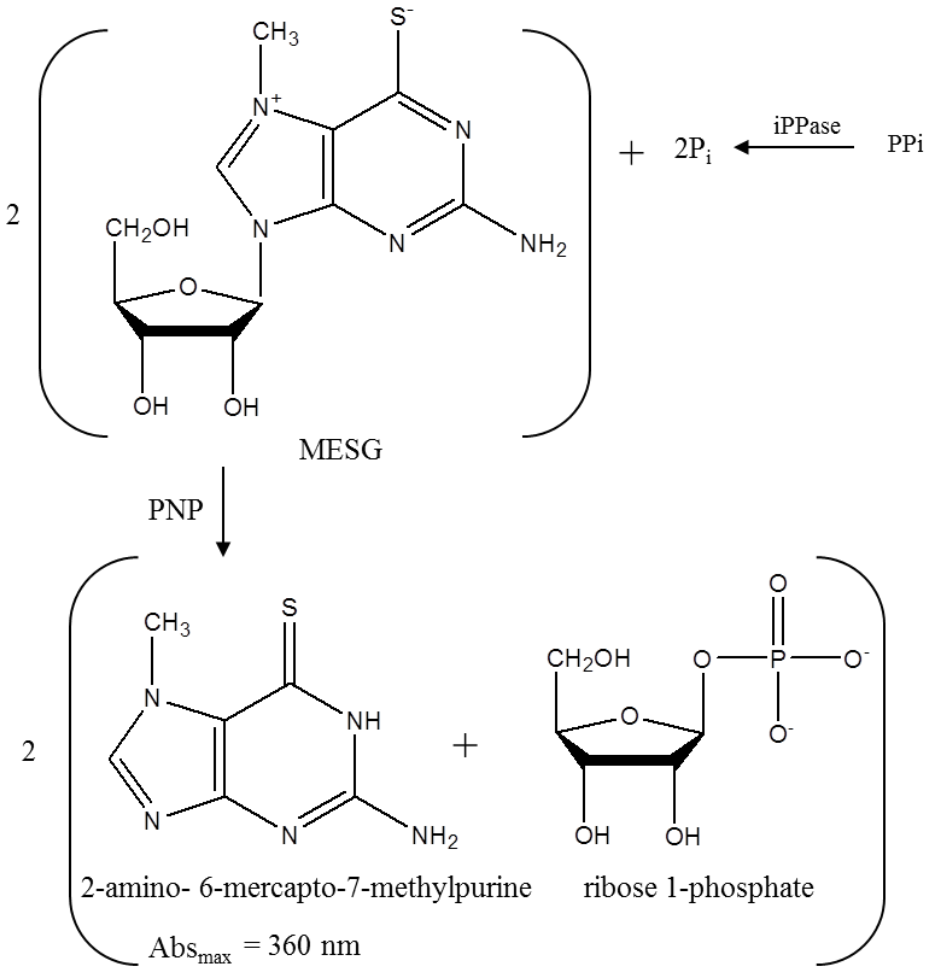


Figure 3. MESG assay for the detection of pyro phosphate

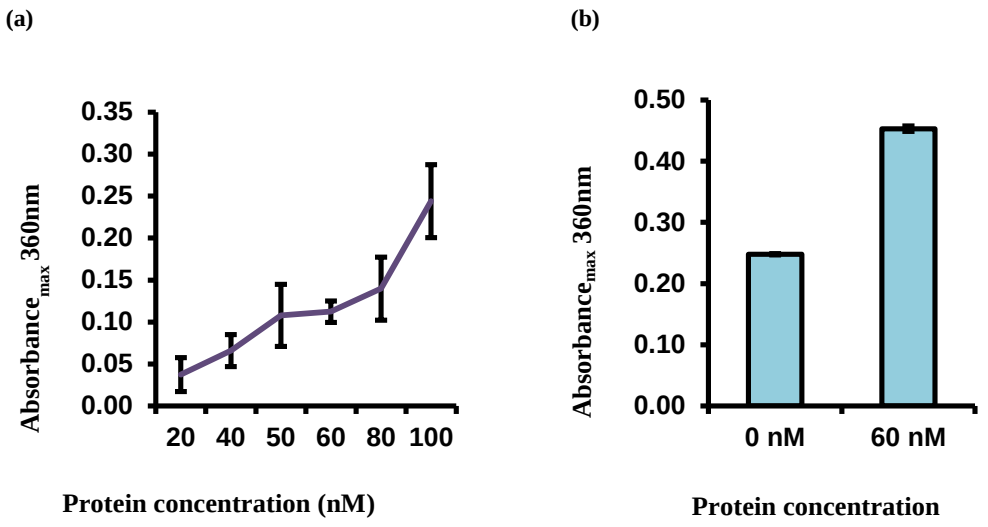


Figure 4. Optimization of IspD enzyme assay. (a) IspD enzyme concentration was varied initial and end point absorbance was recorded at 360 nm. The graph represents the variation of endpoint absorbance as a mean ±SD for three separate experiments. (b) Assay performed with 0.06 U/ml PNP showed maximum spread in absorbance values between the reactions performed with and without IspD and is represented as mean ±SD for three separate determinations.

The final assay was performed in 50 mM Tris pH 7.5 supplemented with 1 mM $MgCl_2$. 60 nM ISPD, 0.1 U/ml IPPase, 0.06 U/ml PNP and 200 µM MESG was used. 1 µl of DMSO was added as a control since all the compound libraries were dissolved in DMSO. The reaction was initiated by the addition of the substrates (50 µM MEP and 100 µM CTP). The total reaction volume was 50 µl. The reaction was monitored by reading the absorbance at 360 nm.

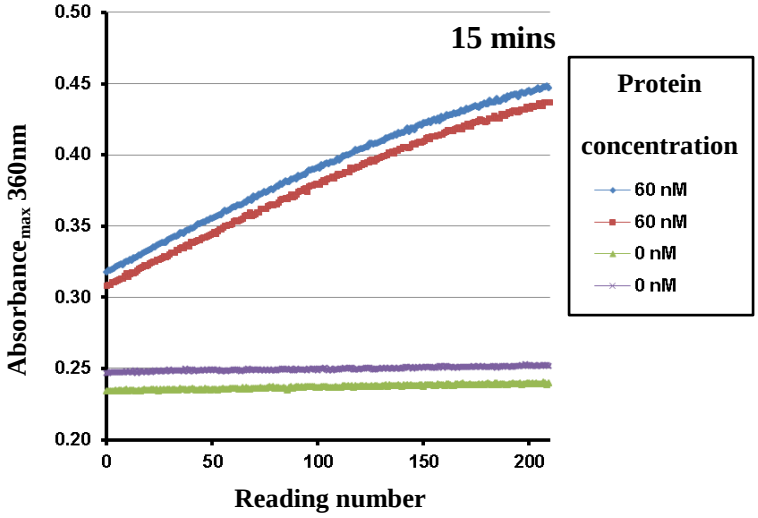


Figure5. IspD enzyme activity monitored over a 15 minute interval. The experiments were done in triplicates and repeated twice. Two of the representative curves are shown above.

2.2 The adaptation of the MESG assay for IspD to a 384-well format

In order to test the response of the assay developed, it was performed in a 384 well plate by filling the entire plate with 4 different reaction conditions: the optimized assay, assay without IspD, assay without substrate and assay without dye. These were distributed on the plate so as to account for any positional effects that may exist. None were observed and it was noticed that the assay without IspD/substrate generated the least signal. Since there were no known inhibitors of M.tb IspD and the available herbicide inhibitors of *Arabidopsis thaliana* IspD did not show inhibition when tested, assay without substrate was used as positive control. (This was also convenient for CyBi Robot programming compared to the assay without IspD).

Subsequently, a program was written for the CyBi®-Well vario and a screen was performed against a randomly selected plate in the SAC1 diversity library. The final compound concentration was 20µM. The compounds were incubated for 5 minutes which was a compromise between MESG dye decay and CTP hydrolysis rates in real time and the need to pick as many different kinds of lead molecules as was possible through the screen. The program was further optimized to give the result shown below (Figure 6).

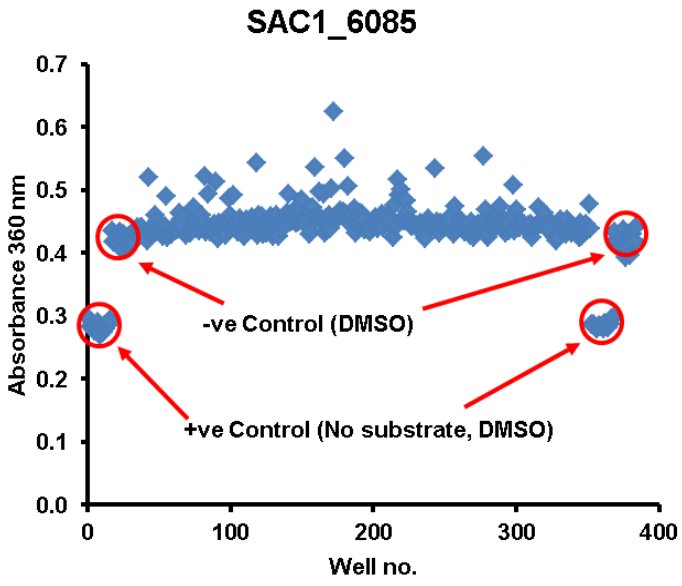


Figure 6. Scatter plot representing IspD screened against SAC1 diversity library plate no. 6085

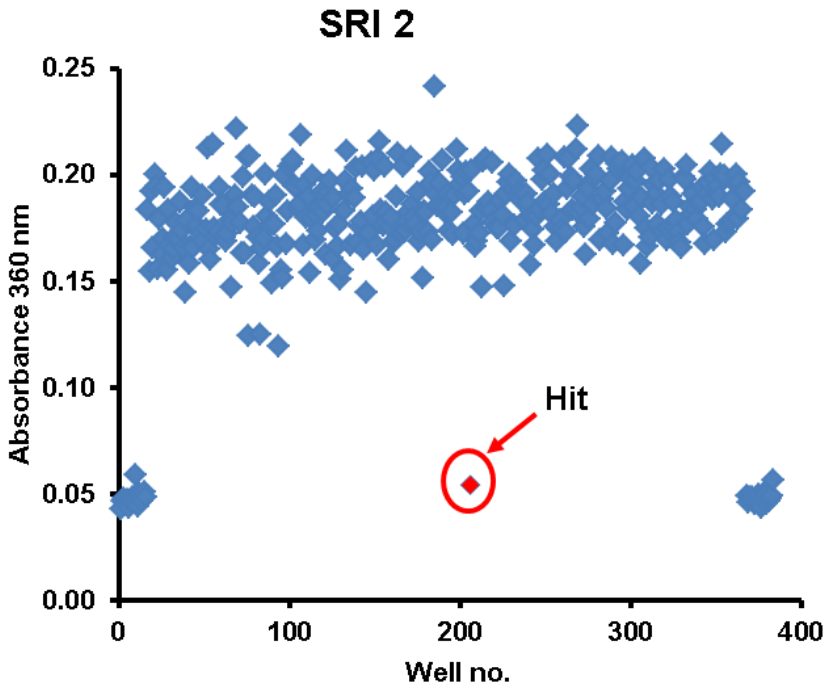


Figure 7. Scatter plot representing IspD screened against SRI library plate 2

Based on the above results, it was decided that the initial time=zero absorbance would be subtracted from all the absorbance readings in order to have the baseline at zero. Since the spread of the absorbance values were within ±3x standard deviation on average (though definitely skewed in this case toward the higher values), it was determined that 20 µM was a fitting concentration to perform the screen. For the initial library, the hit rate would be defined as Hit = Mean-(3x standard deviation).

The first library to be screened was the SRI library which was developed by the Southern Research Institute. All the compounds in this library are active against the H37Rv strain of M. *tuberculosis*. The library is enriched for kinase inhibitors, which

reduces the available diversity to an extent [38]. The result from a typical plate is shown above (Figure 7).

2.3 The screening results

1241 compounds were screened and the hits as characterized by the previously established definition were retested at 40, 20 and 10 µM in triplicate. 14 of the original hits could be re-confirmed as true hits. These were subsequently screened for IPPase and PNP inhibition using the secondary screen. 12 of the hits dropped out at this stage and 2 compounds could be confirmed as true IspD inhibitors. Since this was the first library screen performed, 10 compounds (which were not hits in the original screen) were selected from each plate and re-tested in duplicate at 20 µM concentration to identify the incidence of false negatives. None of the selected compounds showed inhibition. It was then decided to proceed with all future libraries using the same strategy.

The next library to be screened was a section of the SAC1 50,000 compound diversity library. This section was comprised entirely of ChemBridge Hit2Lead® supplied compounds. Not all compounds in this selection were whole cell active; nevertheless, they were easy to procure for further studies. The hits from the SRI library had presented some challenges on this front. However, this was a temporary deviation from the established program and the whole cell activity for all of these compounds was available in our lab. 2112 compounds were screened, 20 re-tested and 3 of these were confirmed as true IspD inhibitors.

The Sac1 whole cell hit library was screened next. 1667 compounds were screened, 14 re-tested and 3 were true IspD inhibitors. 2 of these overlapped with the SAC1 ChemBridge Hit2Lead® library screen hits.

Three more whole cell active libraries were tested: the Natural Product library from Malaysia (145 compounds), the Herbicide library (52 compounds) and the NIH clinical collection (445 compounds). None of these libraries yielded any hits. The six hit compounds from the various screens are shown below (Figure 8).

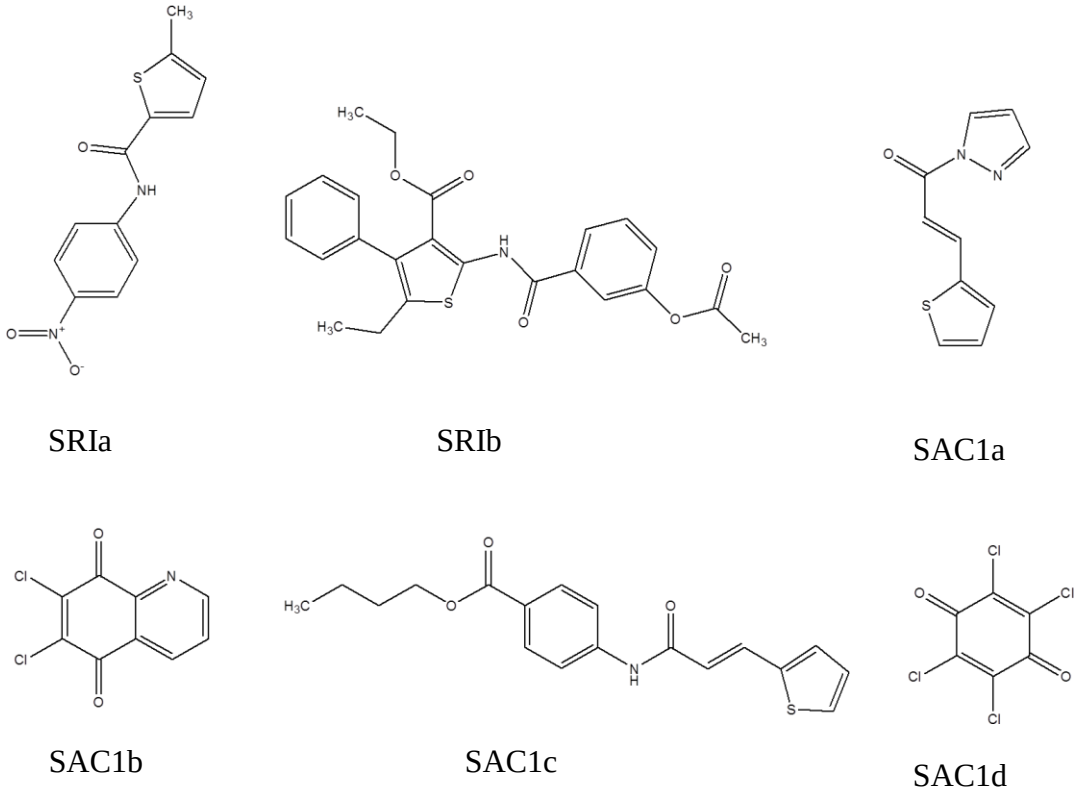
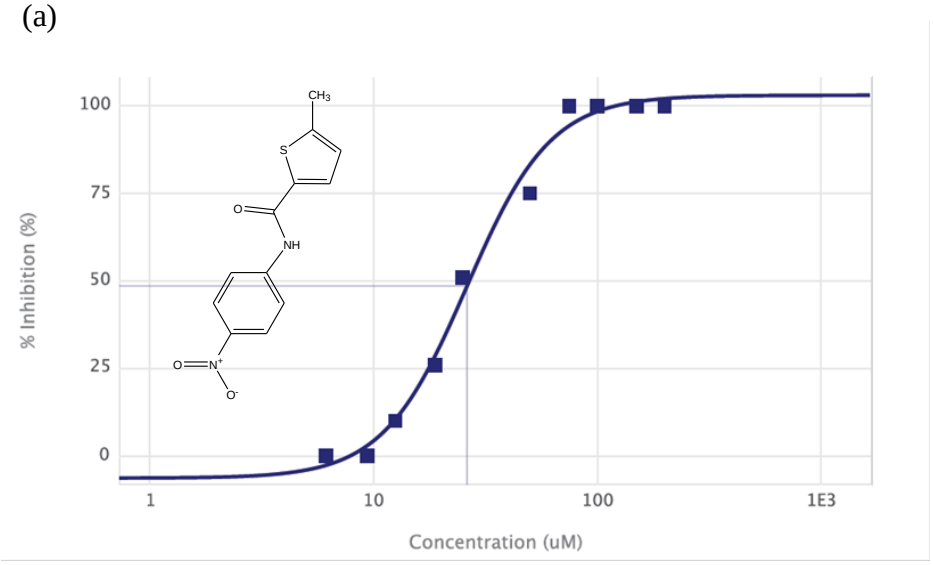


Figure 8. Final hits from the IspD enzyme screen

2.4 Determination of IC_{50}

The IC_{50} was determined for the hits validated. All the compounds were purchased from ChemBridge Corporation, USA, save one SRI library hit molecule (SRIa). The latter was synthesized in our lab by Dr. Vijay Gawandi. Dose response curves were plotted and IC_{50} calculated using the data fitting algorithm on CDD (Figure 9a, 9b, 9c and 9d). SAC1 d had an IC_{50} that was above 50 µM and since it was similar to SAC1b, is not shown here. In case of SRIb, the fresh compound did not re-test and mass spectroscopy suggests that the compound may have been degraded in the screening plate. Attempts to isolate the active species were unsuccessful.

(a)

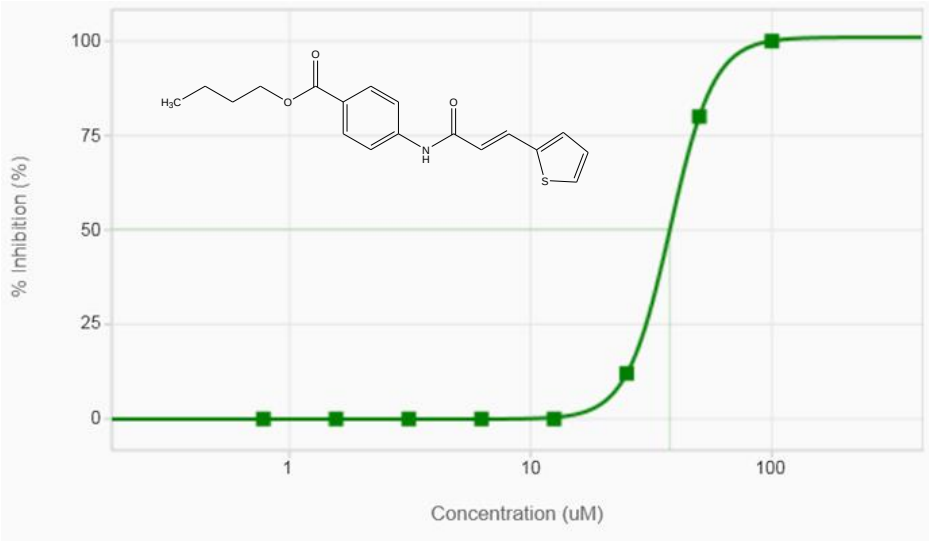


SRIa

IC_{50} = 26.1 µM

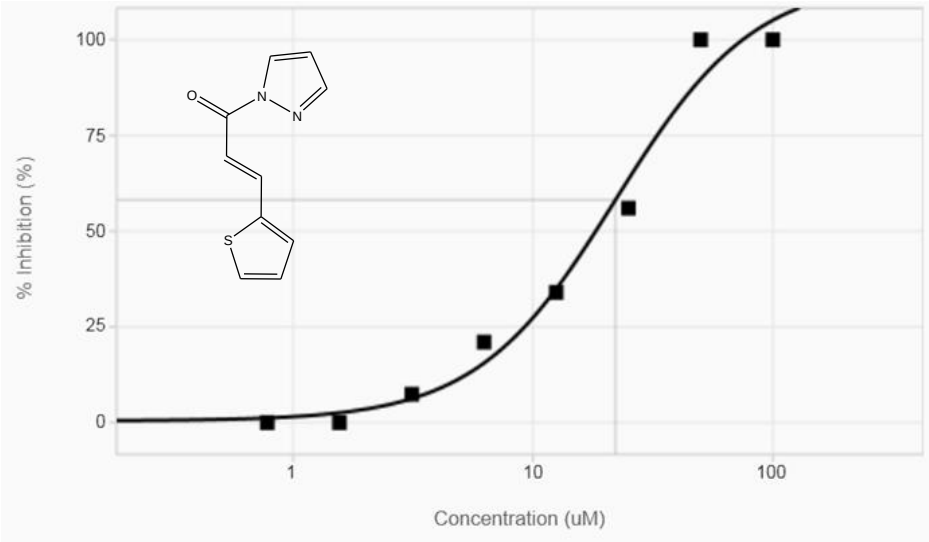
Figure 9. Dose response curve for (a) SRIa (b) SAC1c (c) SAC1a (d) SAC1b. Dose response was performed in triplicates at 10 different concentrations. The curves were calculated using mean over three separate determinations.

(b)



SAC1c

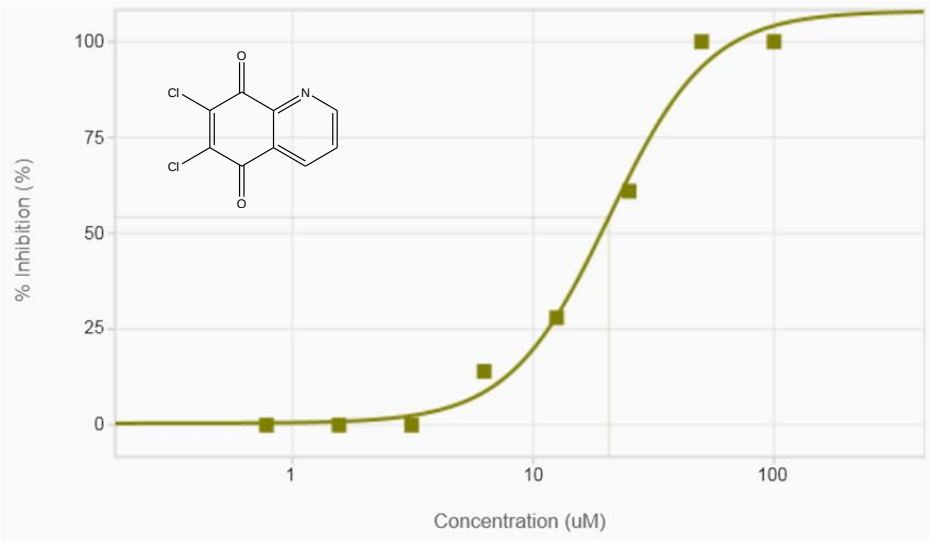
IC_{50} = 37.8 μM



SAC1a

IC_{50} = 22.1 μM

Figure 9. Continued



SAC1b

IC_{50} = 20.6 μM

Figure 9. Continued

2.5 Whole cell assays

The hits were screened against *M. tuberculosis* $MC^{2}7000$ cells and all the compounds were found to be active except SAC1a (confirming previous results from our lab). It was also observed that SAC1b could reduce cell viability at concentrations several fold lower than its IC_{50}. This seemed to indicate that IspD may not be its main cellular target (Figure 10).

The MICs for SRIa and SAC1c however were consistent with the dose response data and we decided to give these compounds top priority.

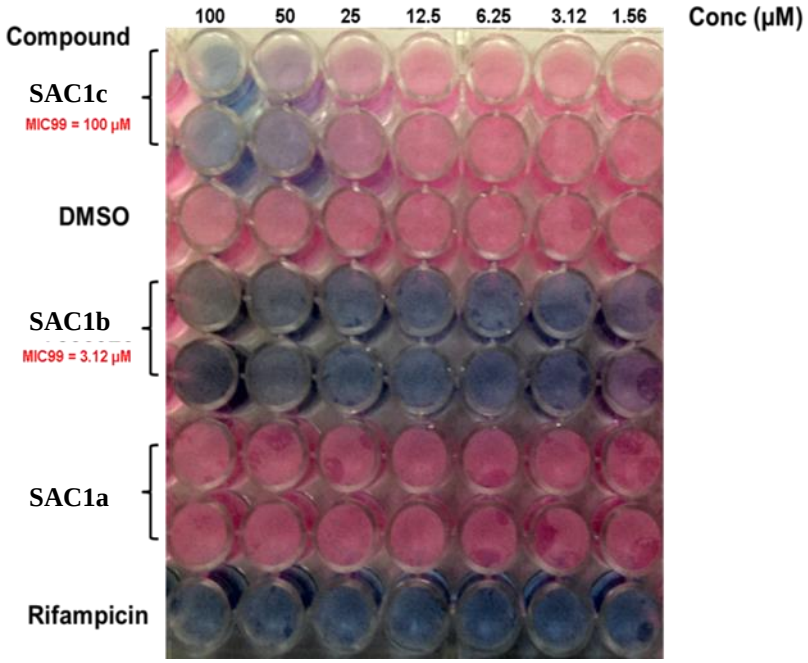


Figure 10. MC^{2}7000 growth inhibition by IspD screen hits. Whole cell assays were performed in duplicates and repeated twice. Varying concentrations (100-1.56 µM) of compound were used. A representative plate is shown above. MIC_{99} was calculated as average of two separate determinations.

The compounds were also screened against the human primary HDF cell line (Human Dermal Fibroblasts) using the Resazurine assay to assess cytoxicity. Mitoxantrone was used as a positive control while DMSO was used as a negative control. Mitoxantrone is routinely used in the lab for this purpose and kills cells completely at less than 100 µM concentration. The completion of the assay revealed that SAC1b had an MIC_{90} of 6.25 µM. This established that SAC1b was generally cytotoxic and was subsequently eliminated from the list of hits to be further characterized. (Figure 11)

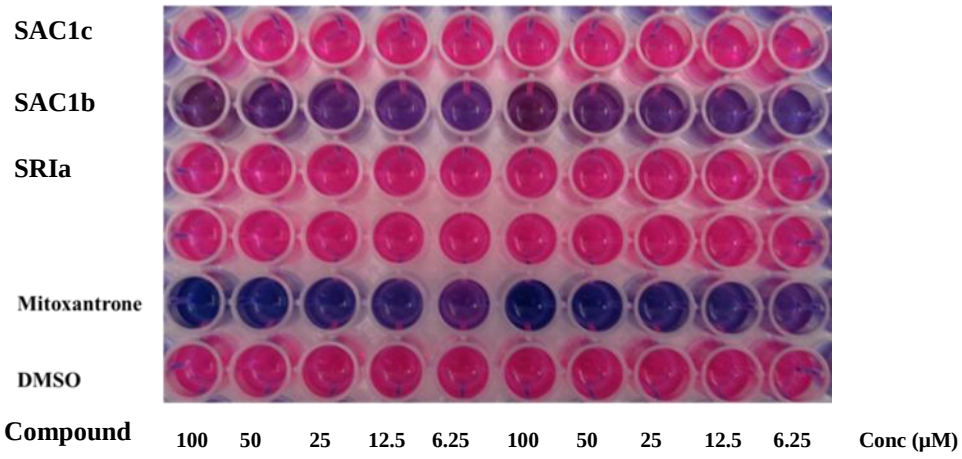


Figure 11. Human dermal fibroblasts growth inhibition by IspD screen hits. Cytotoxicity was determined by staining with resazurin. The experiments were performed in duplicates once at varying concentrations (100-6.25 µM).

2.6 Exploring chemical space with structural analogs

A structure search was performed on SciFinder® using an 85% similarity threshold. 14 hits were obtained for SAC1a with 5 of these available commercially (Figure 12a, 12b, 12c, 12d and 12e). The enzyme assay was performed against all 5 compounds and none of them gave better activity than the initial hit, SAC1a. Compounds 12b and 12c showed minimal activity at 100 µM and 12e had an IC_{50}> 50µM. Compound 12a exhibited 25% of the inhibition shown by the original compound, SAC1a, at the same concentration while 12d had the same activity and dose profile as the original compound. This seems to indicate that substitution at the ortho-position on the benzene ring is not favored. It also suggests that the length of the chain at the meta-position on this ring does not influence binding.

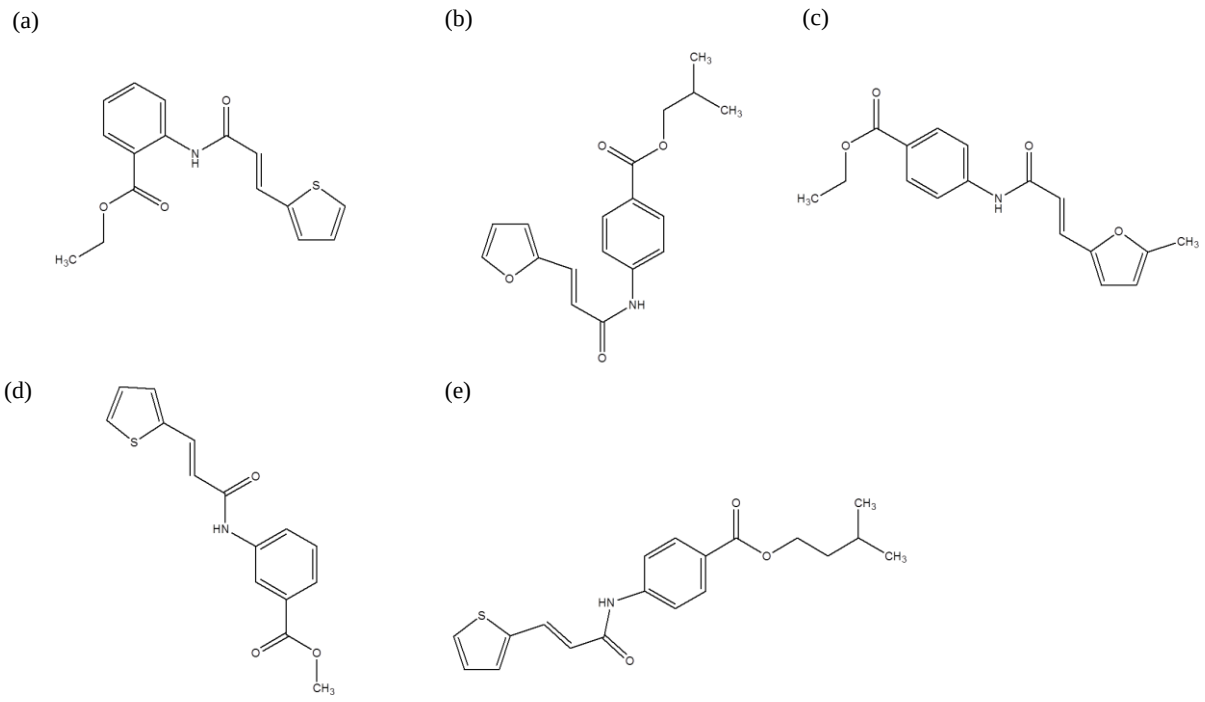


Figure 12. Hits obtained using structure similarity search on Scifinder.

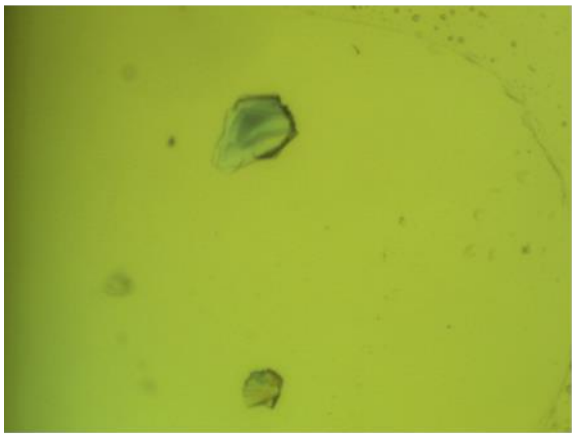
2.7 Crystallization trials

Attempts were made to obtain native, CTP-bound and protein-inhibitor co-crystals. Native and CTP co-crystals were set up in the original conditions optimized in our lab in sitting drop, hanging-drop and micro-batch experimental set-ups (Figure 13d). The initial crystals which diffracted very poorly were crushed and used as seeds for future experiments. Attempts were also made to soak some of the older crystals set up by various members of our lab. These were densely layered and difficult to separate. The few layers that could be isolated diffracted very poorly (4 Å at Advanced Photon Source, Chicago).

IspD was incubated with the inhibitors and screened against the 8 crystallization screens available in the lab. This was done (a) after a single-step nickel (Ni) column purification, (b) after 2-step purification with Ni column followed by gel filtration and (c) after TEV cleavage with single step and two-step purification. Multiple protein concentrations, compound concentrations and protein: reservoir ratios were used to set up plates. After 2 months, a hit was obtained in the condition which has 0.1 M imidazole and 1 M sodium citrate. This was non-cleaved protein that had been subjected to the 2 step purification and concentrated to 8 mg/ml and incubated with 5 mM of compound. This crystal diffracted to 3.8 Å (Figure 13a). Efforts were made to improve the crystal by using the detergent and additive screens. Using the detergent C_8E_5 improved the crystal quality slightly and it could now diffract up to 3.2-3.4 Å (Figure 13b) and it now appeared in little more than a fortnight. Further attempts to improve the quality using the ionic liquid screen, dehydration, annealing and cross-linking have not succeeded so far.

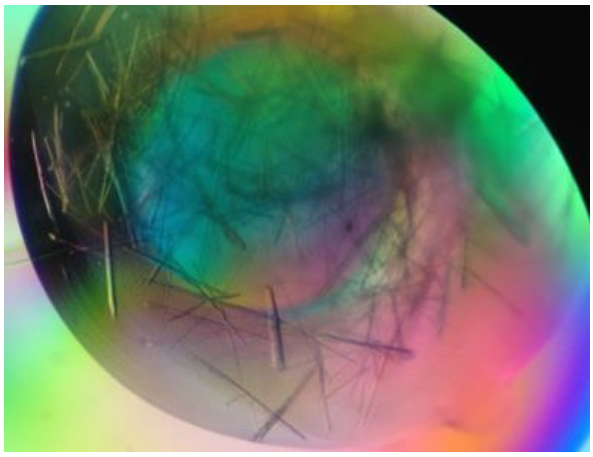
Attempts were also made to soak some of the CTP co-crystals we had managed to reproduce recently (Figure 13c). They were extremely fragile and did not respond well to being soaked in the mother liquor, DMSO or 1-50% DMSO in mother liquor. Different solvents like acetonitrile, isopropanol and ethanol were also used, solo or in combination with mother liquor. The crystals disintegrated even with 3-5 minute soak times.

(a)



(b)

(c)



(d)

Figure 13. IspD crystals and diffraction patterns. (a) IspD-SRIa co-crystal, (b) IspD-SRIa diffraction pattern (3 min exposure), (c) IspD-CTP co-crystals and (d) IspD apo-crystal.

CHAPTER III

PURIFICATION AND CHARACTERIZATION OF Wag31:

RESULTS

3.1 Purification

The *wag31* gene from M. tb was amplified and cloned into pET28 (b) TEV vector and transformed into the *E.coli* BLD1 (DE3) protein expression strain. Expression tests were conducted at multiple temperatures (37°C, 24°C and 16°C). Maximum expression was observed after induction at 37°C. Small scale expression tests determined that the protein was in the soluble fraction (Figure 14).

Although multiple attempts were made to purify the protein, every attempt to lyse it immediately resulted in the formation of a very viscous solution which solidified to a jelly-like consistency on spinning in the high speed centrifuge.

The use of 6 different detergents did not resolve the consistency issue. A typical lysate and supernatant sample are shown below (Figure 14). Almost all the protein is in the soluble fraction. Purification using acidic or basic buffers and high temperature purification (37°C, 42°C) was attempted next. The final purification strategy involved lysing the cells using a sonicator, heating the lysate at 42°C for 5 minutes followed by spinning down in the high speed centrifuge at 16000 rpm for an hour. The supernatant was loaded on a Ni column and purified by the application of an imidazole gradient (0-

500 mM imidazole) (Figure 15). Western blot analysis confirmed the protein was hexa-His Wag31 (Figure 16).

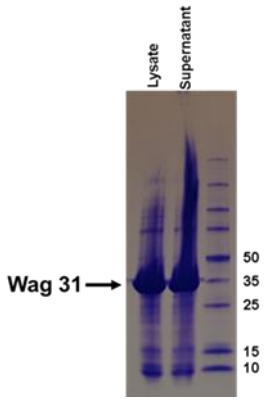


Figure 14. SDS gel seperation of lysate and supernatant fractions of Wag 31 overexpressed in *E.coli* BL21 (DE3) cells. The gel is representative of three separate experiments.

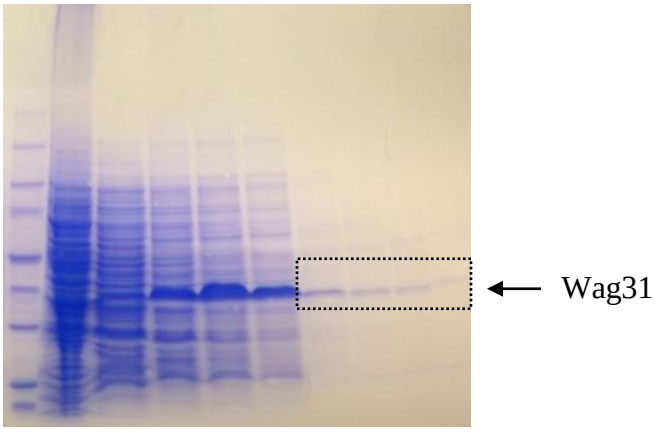


Figure 15. SDS gel separation of Ni affinity chromatography wash and gradient fractions obtained on purifying Wag31. The pooled fractions are enclosed in the box. The gel image is representative of three separate experiments.

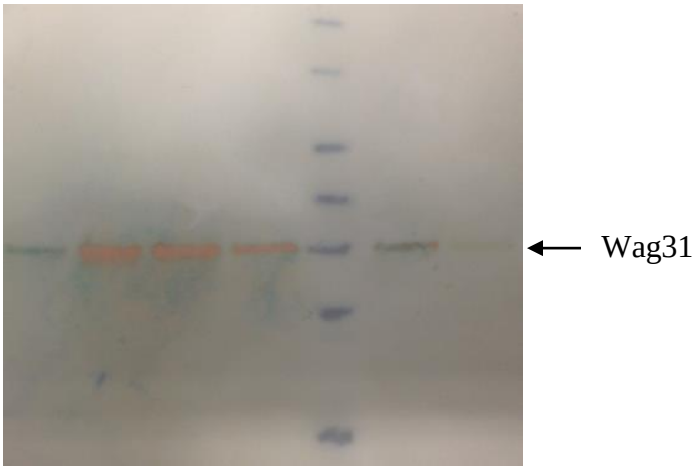


Figure 16. Western blot analysis of Wag31 Ni chromatography fractions probed with anti-His antibody from mouse conjugated with alkaline phosphatase. Substrate was added to trigger the colorimetric reaction. Image is representative of two separate experiments.

Several different constructs of wag31 were purified after the process was optimized. The results are summarized Table 1:

	Construct	**Purification Result**
1	Wag31 WT pet28(b)TEV	Can be purified, forms gel on concentration
2	Wag31T73A(dephosphorylated) pet28(b)TEV	Can be purified, forms gel on concentration
3	Wag31T73E(phosphorylated) pet28(b)TEV	Can be purified, forms gel on concentration
4	Wag31 WT (expressed in *M. smegmatis*) p1602-dest 17	Can be purified, forms gel on concentration
5	Wag31 *M. smegmatis* ortholog MCSG7	Cannot be loaded on the affinity column, forms a strong thick gel on lysis

Table 1. Summary of Wag31 constructs and the behavior of the corresponding proteins on purification. Representative of three separate purifications.

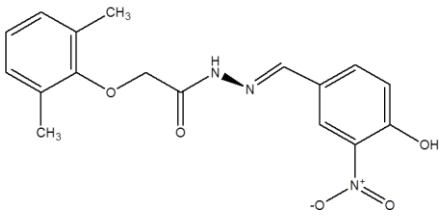
	Construct	Purification Result
6	Wag31 9-96 aa (C truncation) MCSG7	Can be purified, forms gel on concentration
7	Wag31 97-260 aa (N truncation) MCSG7	Can be purified, Most of the protein remains in gel fraction
8	Wag31 128-260 aa (N truncation) MCSG7	Can be purified, Most of the protein remains in gel fraction. Pure protein spontaneously forms gel

Table 1. Continued

3.2 Resistant mutant studies

During the course of this work, a parallel study in the lab led to the isolation of MC^{2}7000 (*M. tuberculosis* BCG vaccine strain) mutants which were resistant to 2 compounds in the SRI compound library (Figure 17). Subsequent whole genome sequencing established that these mutations were in the wag31 protein coding region.

(a)



(b)

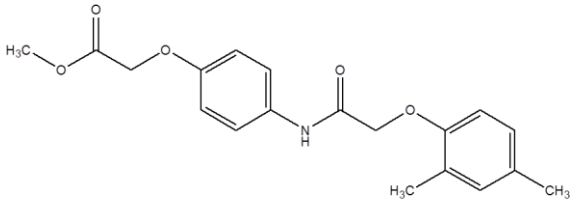


Figure 17. Compounds that led to the isolation of MC^{2}7000 resistant mutants with point mutations in the Wag31 protein coding region (a) Compound A and (b) Compound B

The mutations identified in the initial study were all point mutations in the C- terminal coiled-coil domain of the protein. These include: I194L, E206G and R202C (Data Courtesy: Dr. Inna Krieger and Dr. Thomas Ioerger). Furthermore, a similar study on *M. smegmatis* yielded a mutant resistant to compound A. Whole genome sequencing later established this mutation was equivalent to the E206G mutation in M.tb (In this case E216G) (Data courtesy: Ryan Hughes). 13 more $MC^{2}7000$ mutants resistant to compound A were isolated. Genomic DNA was extracted from these mutants and sent for whole genome sequencing. All of the colonies isolated, had point mutations in the C terminal coiled-coil domain of the wag31 protein coding sequence (Table 2).

Colony No.	Mutations
1	*I194L*
2	K179E, *I194L*
3	*I194L*
4	Q201R
5	*E206G*
6	*I194L*
7	*E206G*
8	D186G
9	Q201R, D181N
10	*E206G*
11	*E206G*
12	K190N
13	*I194L*, R157C, K127E

Table 2. List of $MC^{2}7000$ compound A resistant mutants with the corresponding mutations identified. Thirteen separate colonies were picked to perform genomic sequencing.

The above mutations are mapped in the protein coding sequence in bold red (C-terminal coiled-coil is underlined):

MPLTPADVHN VAFSKPPIGK RGYNEDEVDA FLDLVENELT RLIEENSDLR QRINELDQEL AAGGGAGVTP QATQAIPAYE PEPGKPAPAA VSAGMNEEQA LKAARVLSLA QDTADRLTNT AKAESD***K***MLA DARANAEQIL GEARHTADAT VAEARQ***R***ADA MLADAQSRSE AQLRQAQE***K***A ***D***ALQA***D***AER***K*** HSE***I***MGTINQ ***QR***AVL***E***GRLE QLRTFEREYR TRLKTYLESQ LEELGQRGSA APVDSNADAG GFDQFNRGKN

It was also noted that the E206G and I194L mutations occurred at a higher frequency than the other observed mutations (4 in 10 and 5 in 10 respectively).

3.3 Binding studies

To establish whether Wag31 was indeed the target of compound A, various binding studies were conducted. Intrinsic fluorescence was measured. Since Wag31 does not contain any tryptophans, the signal was quite low. However, there was a reduction in fluorescence in the presence of increasing concentration of compound (Figure 18).

Next, the purified protein was concentrated in the presence of 10 molar excess of compound A. Since the compound was colored (yellow), it was possible to see a clear difference between the compound bound to protein in the concentrator compared to the flow through. Wag31 forms a gel at high protein concentration and it was observed that in the presence of the compound this gel too was colored yellow (Figure 19).

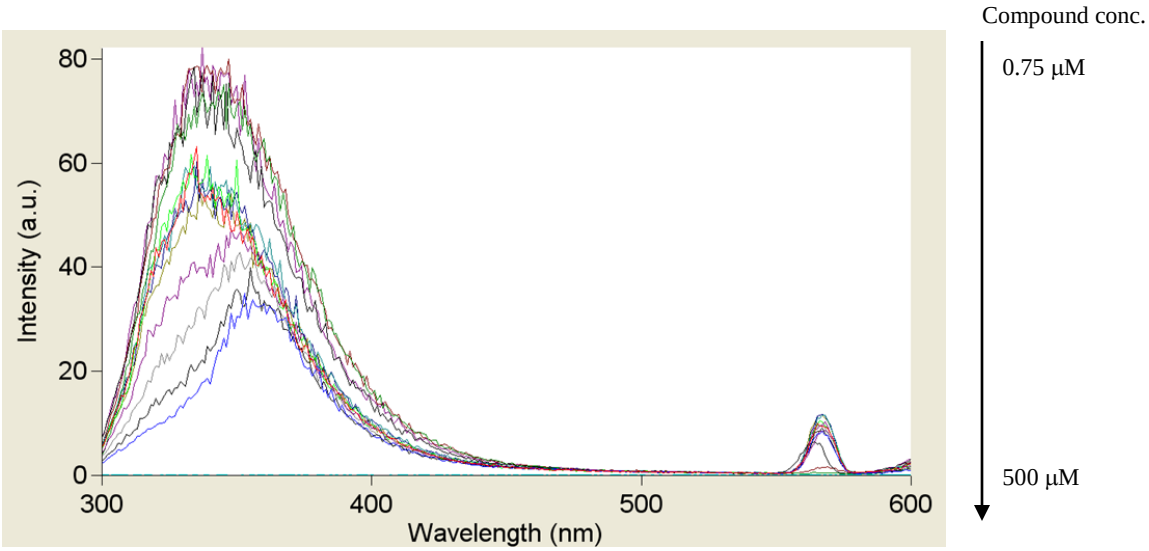


Figure 18. Intrinsic fluorescence measurement of Wag31 (10 μM). Image is representative of three separate experiments.

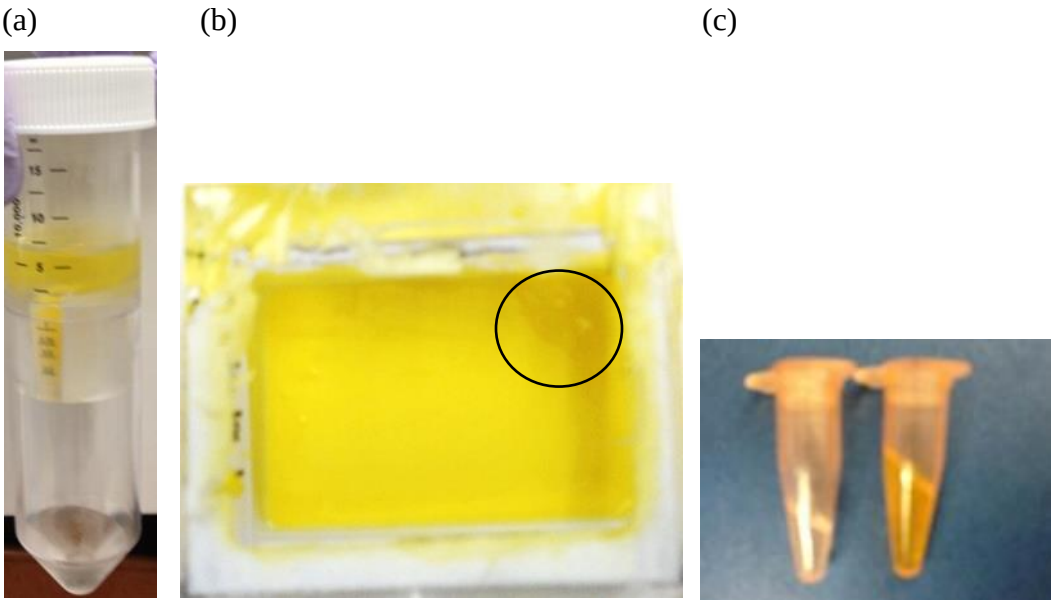


Figure 19. Concentration of Wag31 with 10 M excess of compound A. (a) Concentrator with compound A bound to protein in the upper compartment and flow through in the lower compartment. (b) Gel localized on the concentrator membrane. (c) Comparison of gel formation in the absence and presence of compound A. All the above experiments were performed at least thrice.

3.4 Whole cell viability assays

To further confirm that Wag31 was indeed the target of compounds A and B, a whole cell viability assay was performed using the *M. smegmatis* $MC^2$4517 strain (Figure 20). This strain is routinely used for the overexpression (OE) and purification of TB proteins, including Wag31. If overexpression of Wag31 could induce a shift in the EC_{50} or MIC_{99} values, it would provide further proof that Wag31 was indeed the cellular target of these compounds. Both the wild-type $MC^2$4517 as well as $MC^2$4517 overexpressing *M. tuberculosis* Ribosome Silencing factor (RSF) were used as controls.

It was observed during the course of the experiment that cells over-expressing Wag31 grew at half the rate of the control strains. This was taken into account while reading the plates. Additional readings were taken every day for a week until the values leveled off.

(a)

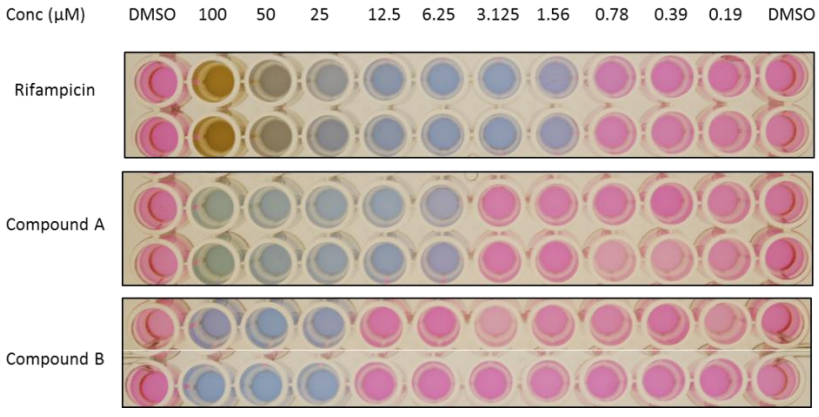
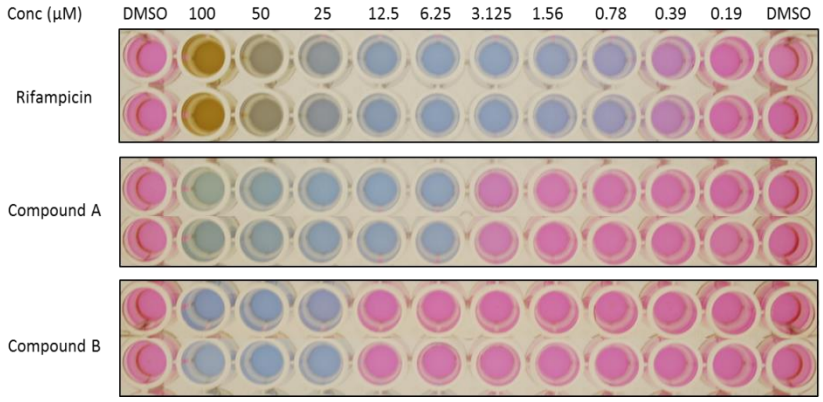


Figure 20. Whole cell growth inhibition assay (a) $MC^2$4517 (b) $MC^2$4517 RSF OE (c) $MC^2$4517 Wag31 OE. The images are representative of two separate experiments.

(b)



(c)

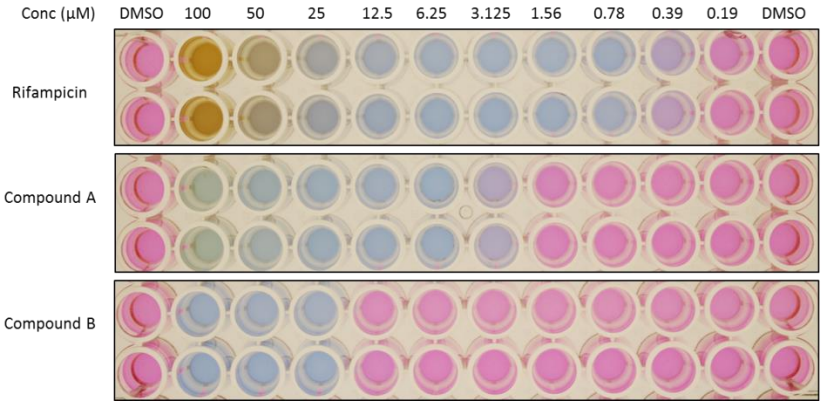


Figure 20. Continued

Compound	Cell Line	EC_{50} (μM)	MIC_{90} (μM)	MIC_{99} (μM)
Compound A	$MC^2$4517 WT	5.51±0.015	6.38±0.020	7.48±0.070
Compound A	$MC^2$4517 Ribosome Silencing Factor OE	5.24±0.015	5.79±0.015	6.45±0.020
Compound A	$MC^2$4517 WT wag31 OE	2.94±0.025	3.41±0.040	4.01±0.060
Compound B	$MC^2$4517 WT	21.8±0.400	24.3±0.350	27.3±0.300
Compound B	$MC^2$4517 Ribosome Silencing Factor OE	22.1±0.650	24.4±0.650	27.3±0.700
Compound B	$MC^2$4517 WT wag31 OE	20.1±0.650	22±0.800	24.2±1.050

Table 3. Summary of EC_{50}, MIC_{90} and MIC_{99} values obtained from whole cell growth inhibition assays. The results shown above are the mean ± SE for two separate experiments

Both compounds showed reduced EC_{50} and MIC values on Wag31 over-expression indicating that Wag31 is the target (Table 3). However, the reduction in values suggests that the drug induces some change in the protein that causes it to be toxic to the cell.

3.5 Gel formation in the presence and absence of compound

To understand whether the compound causes a change in the protein conformational equilibrium, shifting it towards its gelatinous phase, a comparative study was performed. Wag31 WT, Wag31 E206G mutant protein, Wag31 9-96 N terminal construct and Wag31 97-260 C terminal construct were expressed and purified from *E. coli* (3 Liters each).

The quantity of gel in the supernatant was quantified (Figure 21). The C terminal construct was entirely in the gel form and was loaded after diluting 20 fold on a gravity column. However, this did not yield any soluble, non-gelatinous protein. Therefore, we will consider only the first three constructs in this discussion.

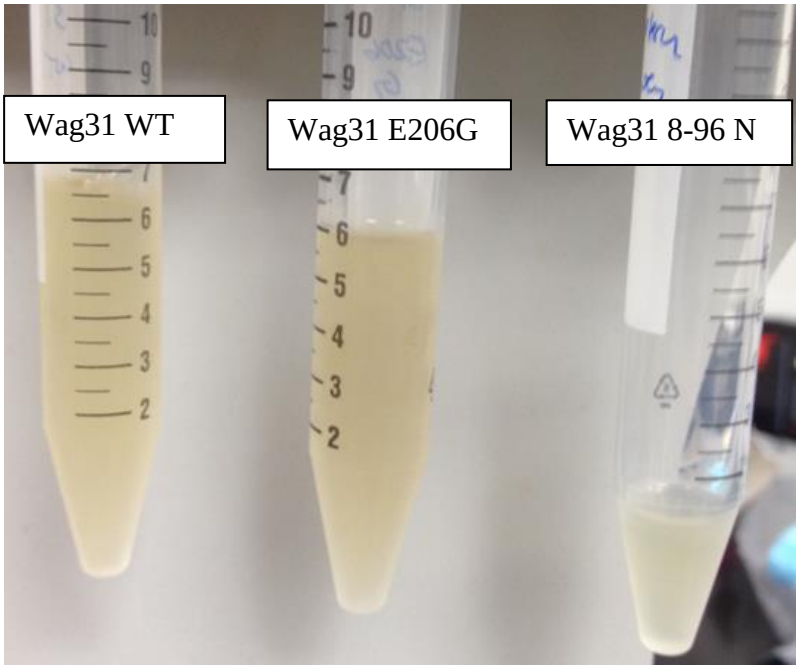


Figure 21. Relative amounts of gel in the supernatant after lysing the cells overexpressing the three Wag31 variants

The purified proteins were then concentrated in the presence and absence of compound (7.5 fold Molar excess). Binding was examined visually by comparing the color

concentration in the concentrated protein (left) versus the flow through (right) (Figure 22).

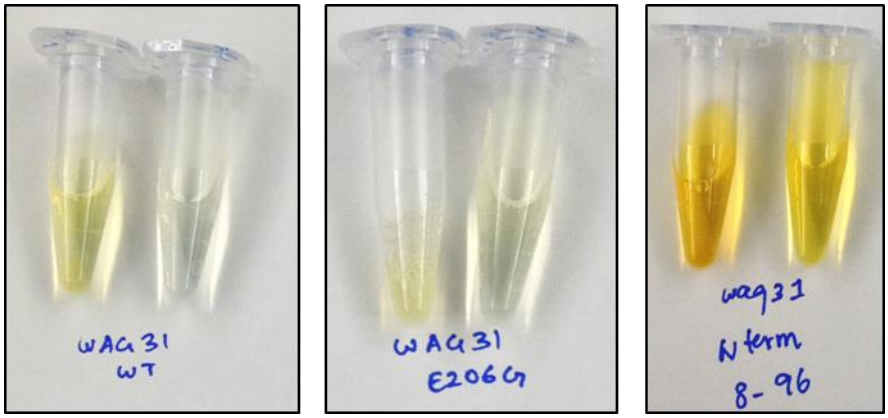


Figure 22. Protein concentrated in the presence of compound A (left) and the respective flow through (right)

While both the WT and N terminal construct seem to bind the compound, the E206G mutant had a visual appearance similar to the flow through even when concentrated down to less than 200 µl. When comparing the gel formation in the presence and absence of compound, a similar pattern was observed (Figure 23).

The compound increased the propensity of the protein to form gels. This was confirmed by measuring protein concentration. In the presence of the compound, a greater proportion of the protein favored gel formation. The gel is promiscuous and has a

tendency to trap most proteins within the mesh. It does not seem to discriminate between Glft2 (M. tb protein) and lysozyme in a pull down assay (Figure 24).

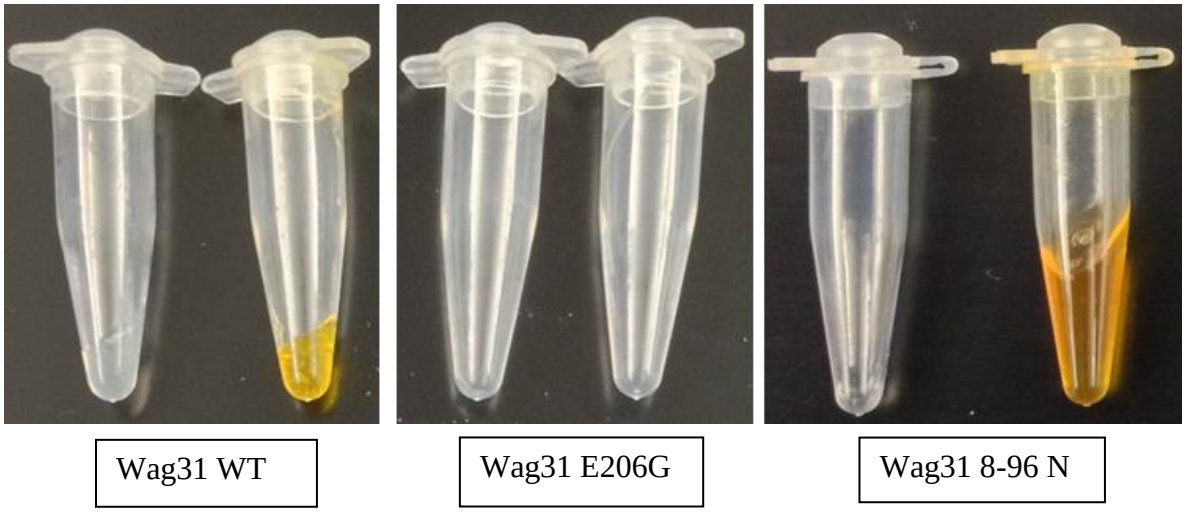


Figure 23. Protein gel formed in the absence (left) and presence (right) of compound A

Since the E206G mutant protein does not seem to form a gel once purified and does not appear to form a gel in the presence of the compound, this may provide a lead concerning the means of resistance in the $MC^{2}7000$ mutants.

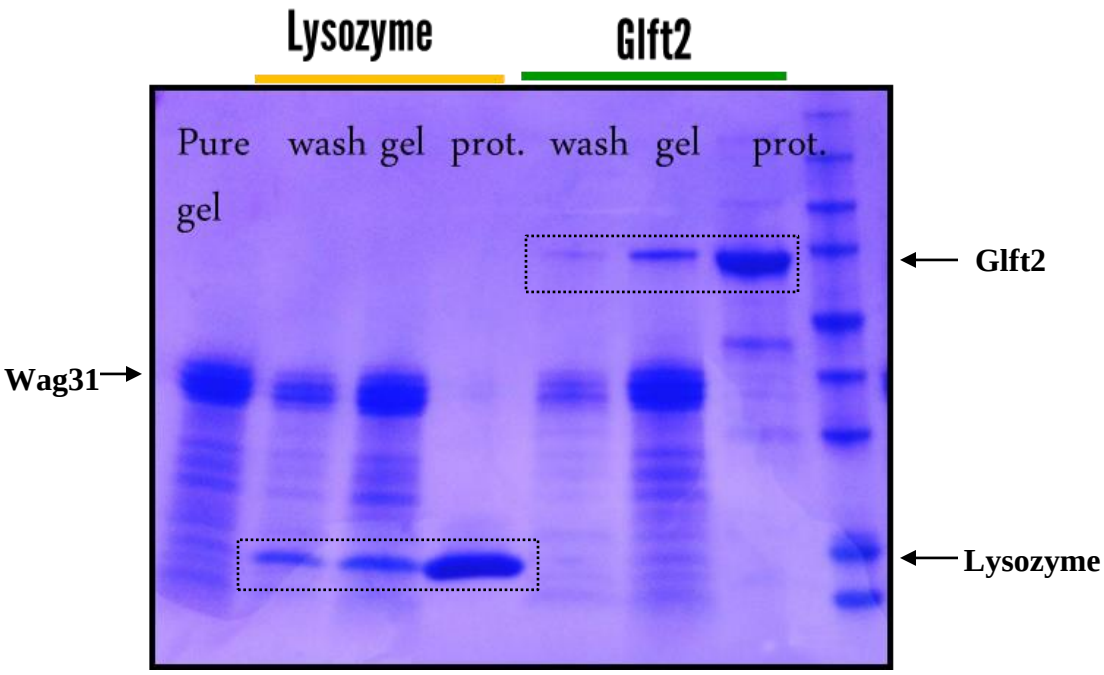


Figure 24. SDS gel separation of pull-down performed using the Wag31 gel as bait

CHAPTER IV

MATERIALS AND METHODS

4.1 Cloning, protein over-expression and purification

The primers used for generating various constructs used in this study are detailed below:

	Construct	Primers
1.	Wag31 WT pet28(b)TEV	F: 5'- GAA TTC CAT ATG CCG CTT ACA CCT GCC -3' R: 5'- CCC AAG CTT CTA GTT TTT GCC CCG GTT GAA TTG ATC -3'
2.	Wag31 WT p1602-dest 17	F: 5'- CGG GGT ACC ATG CCG CTT ACA CCT GCC -3' R: 5'- CCC AAG CTT CTA GTT TTT GCC CCG GTT GAA TTG ATC -3'
3.	Wag31 *M. smegmatis* ortholog MCSG7	F: 5'- TAC TTC CAA TCC AAT GCC ATG CCG CTC ACA CCA G-3' R: 5'-TTA TCC ACT TCC AAT GTC AGT TGT TGC CGC GGT TG-3'
4.	Wag31 9-96 aa MCSG7	F: 5'- TAC TTC CAA TCC AAT GCC CAC AAT GTG GCG TTC AGT AAG CCG -3' R: 5'- TTA TCC ACT TCC AAT GTT AGT TCA TCC CCG CCG AGA C -3'
5.	Wag31 97-260 aa MCSG7	F: 5'- TAC TTC CAA TCC AAT GCC GAG GAA CAG GCC CTG AAG -3' R: 5'-TTA TCC ACT TCC AAT GTT ACT AGT TTT TGC CCC GGT TGA ATT GAT C -3'
6.	Wag31 128-260 aa MCSG7	F: 5'-TAC TTC CAA TCC AAT GCC ATG CTG GCC GAT GCC CAA-3' R: 5'-TTA TCC ACT TCC AAT GTT ACT AGT TTT TGC CCC GGT TGA ATT GAT C -3'

Table 4. List of PCR primers used to generate Wag31 constructs

The primers used for mutagenesis are detailed below:

	Construct	Primers
1.	Wag31T73A pet28(b)TEV	F: 5'-GTT ACG CCG CAG GCC GCG CAG GCA ATC CCG GC-3' R: 5'-GCC GGG ATT GCC TGC GCG GCC TGC GGC GTA AC-3'
2.	Wag31T73E pet28(b)TEV	F: 5'-GTT ACG CCG CAG GCC GAA CAG GCA ATC CCG GC-3' R: 5'-GCC GGG ATT GCC TGT TCG GCC TGC GGC GTA AC-3'
3.	Wag31E206G pet28(b)TEV	F: 5'-CCA GCA GCG CGC GGT GCT TGG CGG CCG CCT CGA GCA GCT GCG-3' R: 5'-CGC AGC TGC TCG AGG CGG CCG CCA AGC ACC GCG CGC TGC TGG-3'

Table 5. List of PCR primers used to generate Wag31 mutants

The gene coding for wag31 was amplified by PCR USING M. tb H37Rv genomic DNA as the template (Table 4). After restriction digestion they were ligated into the pet 28(b) TEV vector (TEV modification courtesy: Dr. Manchi Reddy). In case of the MCSG vectors, we followed the protocol detailed by Eschenfeldt *et al* [39]. The Agilent QuikChange II Site-Directed Mutagenesis Kit was used to generate point mutations. Primers were designed using the PrimerX software (Table 5). The IspD pet 28 (b) vector was generously provided by Cory Thurman. All of the vectors were submitted to GTL to confirm the protein coding sequence.

The different vectors were transformed into BL21 (DE3) protein over-expression strain (Novagen). Overnight culture was prepared and diluted the next morning (1:500) in fresh LB medium containing 50 µg/ml Kanamycin or 100 µg/ml Carbenicillin for pet28 and MCSG7 vectors respectively. The cells were grown in a 37°C shaker. IPTG (isopropyl-β-d-thiogalactopyranoside) was added to a final concentration of 1 mM when the culture reached an optical density at 600 nm (OD600) of 0.8. Induction was carried out for 20 h in an 18°C shaker.

Wag31 WT p1602-dest 17 was transformed into $MC^2$4517 cells. 3 ml cultures were grown at 37°C for a minimum of 2 days before diluting into fresh 7H9 media supplemented with 0.5% Glycerol, 0.05% Tween 80 and 0.5% Dextrose. 80 µg/ml of Hygromycin and 20 µg/ml of Kanamycin were added for selection. Cells were induced by the addition of 20% Acetamide to a final concentration of 0.2% when the O.D. 600 reaches 1.

The cells were re-suspended in a buffer containing 20 mM Tris pH 8.0, 500 mM NaCl and 2mM β-Mercaptoethanol (Tris pH 7.5 with 5% Glycerol in case of ISPD), Protease Inhibitor Cocktail EDTA free set V(Novagen) (500 µl/50ml of lysate), $MgCl_2$ (final concentration 5 mM) and DNase (0.5 mg/ml for IspD, 12.5 mg/ml for wag31). The cells were then disrupted by sonication using a Branson Sonifier 450 (Branson Ultrasonics Corp.) and cell lysates were clarified by centrifugation at 16000 rpm for an hour. The supernatant was filtered using a 0.45 micron syringe filter and applied on a Nickel

affinity column. A step-wise or gradient application of Imidazole (10-500 Mm) yielded 80% pure protein. This protein was then dialyzed overnight against low salt (50-150 mM) buffer. IspD was subjected to an additional purification step before it could be used for enzyme assays. It was loaded on an SE75 gel filtration column and the dimer peak was pooled, concentrated and stored in 50 µl aliquots after flash freezing in liquid Nitrogen. Wag31 could be stored on ice in the refrigerator for 4 days. Proteins were concentrated using Corning Spin-X UF 10 kDa MWCO concentrators (Corning Inc.).

4.2 Protein crystallization trials

Crystal screening was performed using the vapor diffusion method in 96 well sitting drop plates (Intelliplate). The various solutions from the commercial and Lab-designed screens were dispensed and mixed with the protein solution using the Mosquito Crystal Robot (TTP Labtech). Optimization plates were set up using the hanging and sitting drop 26 well plates as well as the micro-batch under-oil plates. For the latter both Al's oil and Mineral oil were used to layer above the crystal drops.

Crystals were screened for diffraction both at the home X-ray Source as well as at beamline 19 ID, Structural Biology Center, Advanced Photon Source (APS), Argonne National Laboratory (ANL).

4.3 Enzyme assays, screening and dose response curves

The IspD enzyme assay was performed in 50 mM Tris pH 7.5 supplemented with 1 mM $MgCl_2$. 60 nM ISPD, 0.1 U/ml IPPase, 0.06 U/ml PNP and 200 µM MESG was used. 1 µl of DMSO was added in the control wells. The reaction was initiated by the addition of the substrates (50 µM MEP and 100 µM CTP). The total reaction volume was 50 µl. The solutions were dispensed using the CyBi®-Well vario liquid handling system (CyBio AG) into Greiner 384-well flat-bottomed plates. The reaction was monitored by reading the absorbance at 360 nm using the Thermo Scientific Multiskan GO plate reader (Thermo Fisher Scientific, Inc.).

The Screening data was plotted on Microsoft Excel and compounds which generated absorbance values below 3X standard deviation were considered hits. Counter screens were performed with the hits in the absence of IspD to eliminate IPPase and PNP inhibitors.

Dose response curves were generated by incubating the enzyme with a serial dilution of the hit compounds (0-200 µM) and then running the enzyme assay. The concentration and % Inhibition were then fit into the Hill plot equation on CDD to calculate IC_{50}.

4.4 Whole cell assays

Whole cell viability assays are performed using Resazurin, a non-fluorescent dye that is converted to pink, fluorescent Resorufin via the reduction reactions that occur in

metabolically active cells. The fluorescence is read using 560EX nm/590EM nm filter settings on the FLUOstar Omega Plate Reader (BMG LABTECH).

The whole cell assays were performed by incubating the cells ($MC^2$7000, $MC^2$4517 and Human Dermal Fibroblast (HDF) cells to evaluate cytotoxicity) with a serial dilution of the relevant compounds AT 37°C, followed by staining with Resazurin. The incubation times varied depending on the growth rate of the particular cell type.

In the wag31 OE assay, the media was supplemented with inducer (acetamide) and antibiotic (Hygromycin and Kanamycin) in order to maintain protein over-expression.
The fluorescence values were used to calculate % inhibition on Microsoft Excel. The concentration and % Inhibition were then fit into the Hill plot equation on CDD to calculate EC_{50}, MIC_{90} and MIC_{99}.

4.5 Intrinsic fluorescence measurement

Intrinsic Fluorescence was measured using 10 µM (final concentration) of protein suspended in 120 µl of 20 mM Tris pH 8.0, 50 mM NaCl buffer. The buffer was used to blank the instrument every 3-4 readings. The compound concentration was increased from 0.75 µM to 500 µM final concentration, doubling the concentration each time. 3 readings were taken for each representative concentration. The fluorescence is read by excitation at 282 nm followed by scanning emission between 280 and 600 nm on the Cary Eclipse Spectrophotometer (Agilent).

4.6 Mutation studies

The resistant mutant colonies were inoculated in fresh 7H9 media and the genomic DNA extracted as detailed by Larsen *et al* [40]. The wag31 protein coding sequence was then amplified by PCR using the wag31 WT primers and cloned into pet 28 (b) TEV vector. These were then sent for sequencing to GTL using the T7 promoter and terminator primers.

The samples were also sent for whole genome sequencing at a later date to identify mutations outside the wasg31 coding sequence.

4.7 Gel harvesting and pull-down assays

The gel formed on concentrating the protein was harvested by cutting out the concentrator membrane with a blade and scraping it with a spatula. The scrapings were stored in PCR tubes on ice for up to 4 days.

For pull-downs, the gel was incubated with the protein of interest in a 1:1 or 1:2 molar ratio for up to 20 minutes at 4°C. It was then washed with 5X buffer (volume) for up to one hour with shaking. The wash buffer was sucked out using a 1ml pipette. The gel was then run on an SDS gel to determine if any protein remained bound. The gel was stained using Coomassie Brilliant Blue.

4.8 Western blot and protein quantification

Protein samples were run on an SDS gel and transferred on to a nitrocellulose membrane. The membrane was the probed using an anti-His antibody peroxidase conjugated (Mouse, Sigma). It was then incubated with peroxidase substrate to generate a colored band on the membrane.

Protein was quantified using Bradford reagent and absorbance at 595 nm read on Carey50 UV/Vis Spectrophotometer. BSA was used as standard.

CHAPTER V

SUMMARY AND CONCLUSION

Mycobacterium tuberculosis presents a significant world health challenge with the emergence of strains resistant to both first and second line drugs. Identifying new lead molecules with alternate targets is the greatest need of this moment. This study details the study of two such targets.

The first study employs a top-down approach, where IspD is first identified as a promising enzyme in an essential pathway. An assay was developed and then adapted to the medium-throughput 384-well format. The enzyme was then screened against various whole cell active molecules which included compounds from our diversity libraries, natural products, NIH approved small molecules and herbicides. Since these compounds are known to be whole-cell active, the challenging issue of breaching the M. tb cell wall is circumvented. Two compounds were identified by the screen with IC_{50}s of 26.1 and 37.8 µM and which were whole cell active at 50 µM concentrations. These could be lead molecules for the rational design of more potent IspD inhibitors.

The study involving Wag31 employed a two-pronged approach. Wag31 was identified as a promising target, being both essential and without a homolog in the host and thus attempts were underway to characterize it and understand its function. Parallel studies conducted within the lab identified Wag31 as the target of one of the whole cell active

compounds in the SRI library. Point mutations within the Wag31 protein coding region of the M. tb and *M. smegmatis* genome conferred resistance to this compound. Furthermore, all the mutations were clustered within the C-terminal coiled-coil domain of the Wag31 protein coding sequence. Increasing the concentration of the compound incubated with Wag31 reduced the protein's intrinsic fluorescence and concentrating in the presence of the compound increased the tendency of the protein to form a gel indicating that the compound interacts with the protein. Wag 31 protein that harbors the same mutation as those identified in the compound A resistant mutants did not form a gel on concentrating in the presence or absence of compound A suggesting that the lack of gel formation may explain the resistance. The over-expression of Wag31 in *M. smegmatis* caused a reduction in EC_{50} and MIC values. This points toward Wag31 being the true target of compound A in the cell and lends further credence to the idea that the compound causes some change in the protein that causes it to be toxic to the cell, i.e., gel formation. This represents a truly unique means of compound action in the Mycobacterial cell and one that must be investigated in detail.

REFERENCES

1. WHO, *2011/2012 Tuberculosis Global Facts*, 2012 Access Date; 19, March, 2014. http://www.who.int/tb/publications/TBHIV_Facts_for_2011.pdf
2. Goldberg, D.E., R.F. Siliciano, and W.R. Jacobs, Jr., *Outwitting evolution: fighting drug-resistant TB, malaria, and HIV.* Cell, 2012. **148**(6): p. 1271-83.
3. *Controlled clinical trial of short-course (6-month) regimens of chemotherapy for treatment of pulmonary tuberculosis.* Lancet, 1972. **1**(7760): p. 1079-85.
4. MSF International*TB Treatment and the Side Effects*, 2010. Access Date; 21, March 2014. http://www.msf.org/
5. WHO, *Frequently asked questions - XDR-TB.*Access Date; 19, March, 2014 http://www.who.int/tb/challenges/xdr/faqs/en/
6. Migliori, G.B., De Iaco, G., Besozzi, G., Centis, R. & Cirillo, D. M. , *First tuberculosis cases in Italy resistant to all tested drugs.* Eurosurveillance, 2007. **12**(20): p. 3194.
7. Velayati, A.A., et al., *Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran.* Chest, 2009. **136**(2): p. 420-5.
8. Udwadia, Z.F., et al., *Totally drug-resistant tuberculosis in India.* Clin Infect Dis, 2012. **54**(4): p. 579-81.

9. Raman, K., K. Yeturu, and N. Chandra, *targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis.* BMC Syst Biol, 2008. **2**: p. 109.

10. Holstein, S.A. and R.J. Hohl, *Isoprenoids: Remarkable diversity of form and function.* Lipids, 2004. **39**(4): p. 293-309.

11. Kaur, D., et al., *Biosynthesis of mycobacterial lipoarabinomannan: role of a branching mannosyltransferase.* Proc Natl Acad Sci U S A, 2006. **103**(37): p. 13664-9.

12. Rohmer M, K.M., Simonin P, Sutter B, Sahm H. , *Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate.* Biochemical Journal 1993. **295**(2): p. 517-524.

13. M., R., *The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants.* . Natural Product Reports, 1999. **16**(5): p. 565-574.

14. Hunter, W.N., *The non-mevalonate pathway of isoprenoid precursor biosynthesis.* J Biol Chem, 2007. **282**(30): p. 21573-7.

15. Grawert, T., et al., *Structure of active IspH enzyme from Escherichia coli provides mechanistic insights into substrate reduction.* Angew Chem Int Ed Engl, 2009. **48**(31): p. 5756-9.

16. Rekittke, I., et al., *Structure of the E-1-hydroxy-2-methyl-but-2-enyl-4-diphosphate synthase (GcpE) from Thermus thermophilus.* FEBS Lett, 2011. **585**(3): p. 447-51.

17. Lee, M., et al., *Biosynthesis of isoprenoids: crystal structure of the [4Fe-4S] cluster protein IspG.* J Mol Biol, 2010. **404**(4): p. 600-10.

18. Henriksson, L.M., et al., *The 1.9 A resolution structure of Mycobacterium tuberculosis 1-deoxy-D-xylulose 5-phosphate reductoisomerase, a potential drug target.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 7): p. 807-13.

19. Shan, S. and X. Chen, *Crystallization and preliminary X-ray analysis of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) from Mycobacterium tuberculosis.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2011. **67**(Pt 7): p. 821-3.

20. Buetow, L., et al., *The structure of Mycobacteria 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase, an essential enzyme, provides a platform for drug discovery.* BMC Struct Biol, 2007. **7**: p. 68.

21. Illarionova, V., et al., *Nonmevalonate terpene biosynthesis enzymes as antiinfective drug targets: substrate synthesis and high-throughput screening methods.* J Org Chem, 2006. **71**(23): p. 8824-34.

22. Brown, A.C., et al., *The nonmevalonate pathway of isoprenoid biosynthesis in Mycobacterium tuberculosis is essential and transcriptionally regulated by Dxs.* J Bacteriol, 2010. **192**(9): p. 2424-33.

23. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Molecular Microbiology, 2003. **48**(1): p. 77-84.

24. Eoh, H., et al., *Characterization of the Mycobacterium tuberculosis 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase: potential for drug development.* J Bacteriol, 2007. **189**(24): p. 8922-7.

25. Richard, S.B., et al., *Kinetic analysis of Escherichia coli 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase, wild type and mutants, reveals roles of active site amino acids.* Biochemistry, 2004. **43**(38): p. 12189-97.

26. Wenjun Shi, J.F., Min Zhang, Xuhui Lai, Shengfeng Xu, Xuelian Zhang, and Honghai Wang, *Biosynthesis of Isoprenoids: Characterization of a Functionally Active Recombinant 2-C-methyl-D-erythritol 4-phosphate Cytidyltransferase (IspD) from Mycobacterium tuberculosis H37Rv.* Journal of Biochemistry and Molecular Biology, 2007. **40**(6): p. 911-920.

27. Bjorkelid, C., et al., *Structural and functional studies of mycobacterial IspD enzymes.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 5): p. 403-14.

28. Ramamurthi, K.S. and R. Losick, *Negative membrane curvature as a cue for subcellular localization of a bacterial protein.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13541-5.

29. Maria A Oliva, S.H., Stefan M Freund, Pavel Dutow,Thomas A Leonard, Dmitry B Veprintsev, Leendert W Hamoen and Jan Lo¨we, *Features critical for membrane binding revealed by DivIVA crystal structure.* The EMBO Journal, 2010. **29**: p. 1988-2001.

30. Lenarcic, R., et al., *Localisation of DivIVA by targeting to negatively curved membranes.* EMBO J, 2009. **28**(15): p. 2272-82.

31. Nguyen, L., et al., *Antigen 84, an effector of pleiomorphism in Mycobacterium smegmatis.* J Bacteriol, 2007. **189**(21): p. 7896-910.

32. Kang, C.M., et al., *The Mycobacterium tuberculosis serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape.* Genes Dev, 2005. **19**(14): p. 1692-704.

33. E. Wolf, P.S.K., and B. Berger, *MultiCoil: a program for predicting two- and three-stranded coiled coils.* Protein Science, 1997. **6**(6): p. 1179–1189.

34. Kang, C.M., et al., *Wag31, a homologue of the cell division protein DivIVA, regulates growth, morphology and polar cell wall synthesis in mycobacteria.* Microbiology, 2008. **154**(Pt 3): p. 725-35.

35. Jani, C., et al., *Regulation of polar peptidoglycan biosynthesis by Wag31 phosphorylation in mycobacteria.* BMC Microbiol, 2010. **10**: p. 327.

36. Dahl, J.L. and D. Lau Bonilla, *The wag31 gene of Mycobacterium tuberculosis is positively regulated by the stringent response.* FEMS Microbiol Lett, 2011. **319**(2): p. 153-9.

37. Mukherjee, P., et al., *Novel role of Wag31 in protection of mycobacteria under oxidative stress.* Mol Microbiol, 2009. **73**(1): p. 103-19.

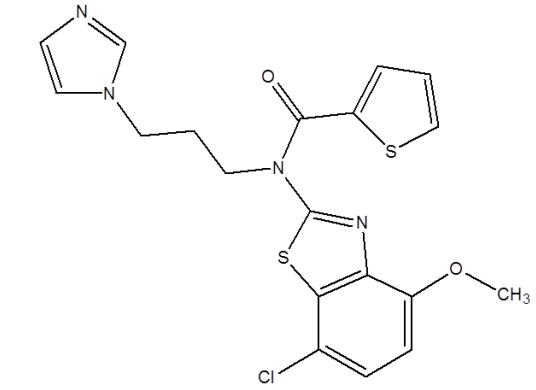
38. Reynolds, R.C., et al., *High throughput screening of a library based on kinase inhibitor scaffolds against Mycobacterium tuberculosis H37Rv.* Tuberculosis (Edinb), 2012. **92**(1): p. 72-83.

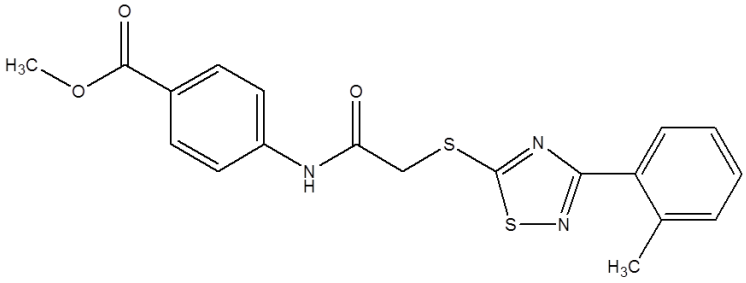
39. Eschenfeldt, W.H., et al., *A family of LIC vectors for high-throughput cloning and purification of proteins.* Methods Mol Biol, 2009. **498**: p. 105-15.

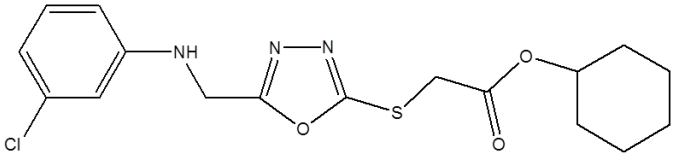
40. Larsen, M.H., K. Biermann, and W.R. Jacobs, Jr., *Laboratory maintenance of Mycobacterium tuberculosis.* Curr Protoc Microbiol, 2007. **Chapter 10**: p. Unit 10A 1.

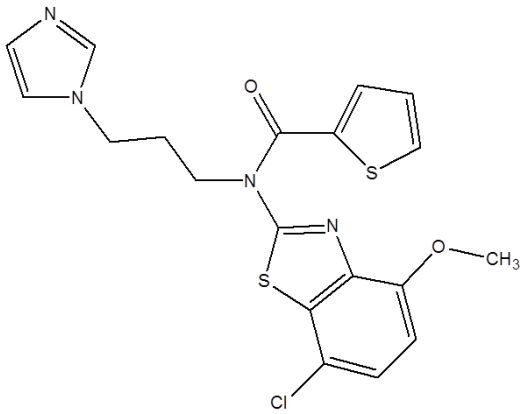
APPENDIX

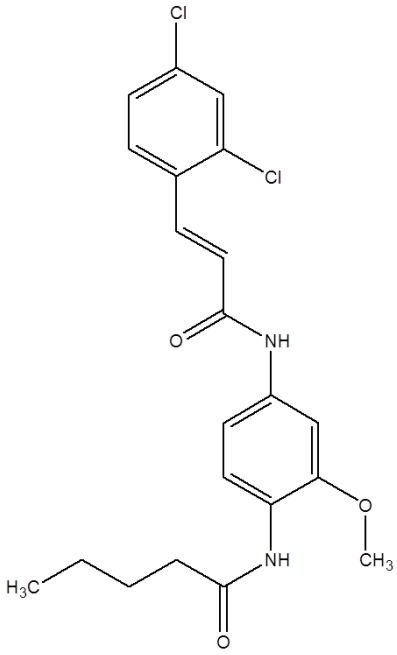
Compounds that dropped out during the secondary screen against PNP and IPPase (IspD project)

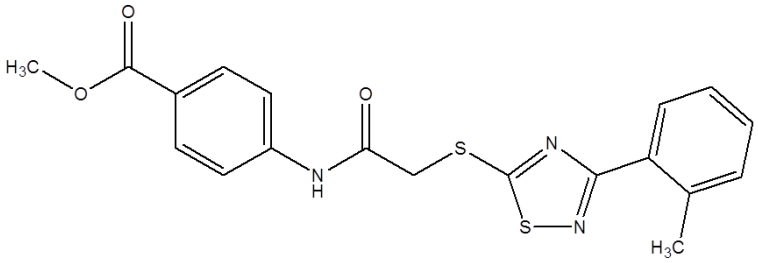


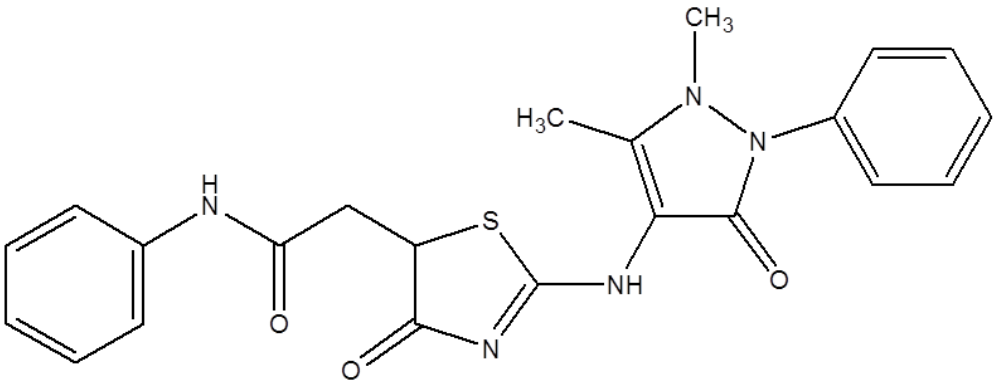


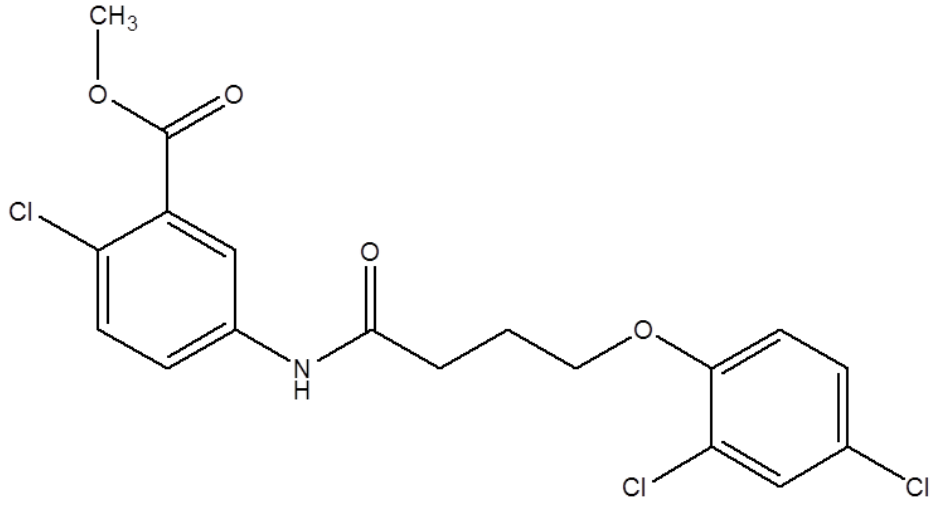


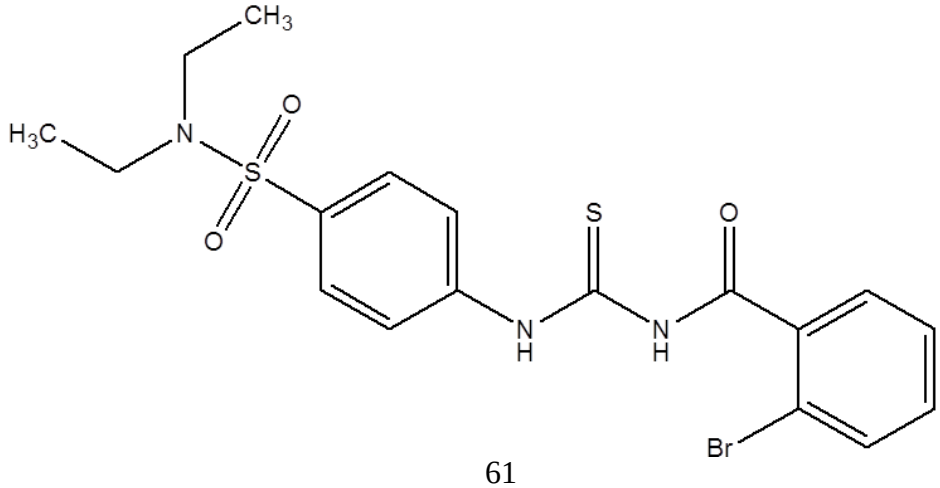


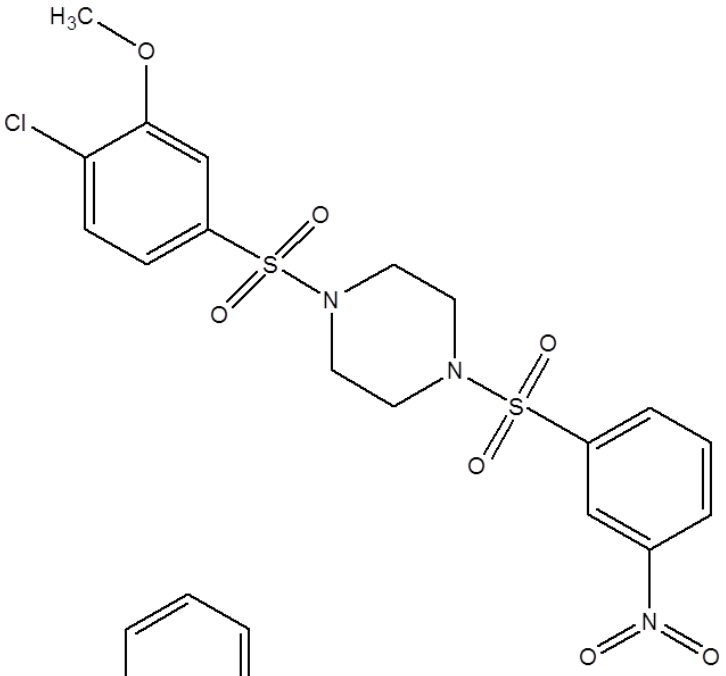


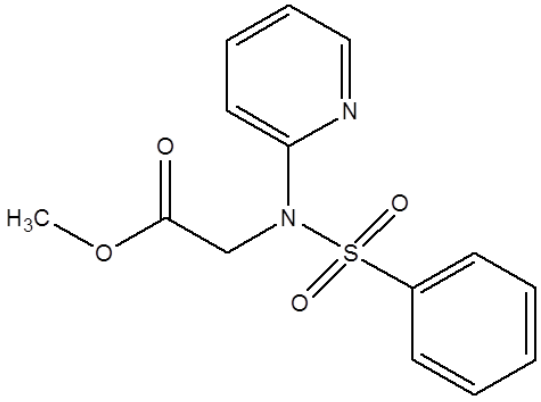


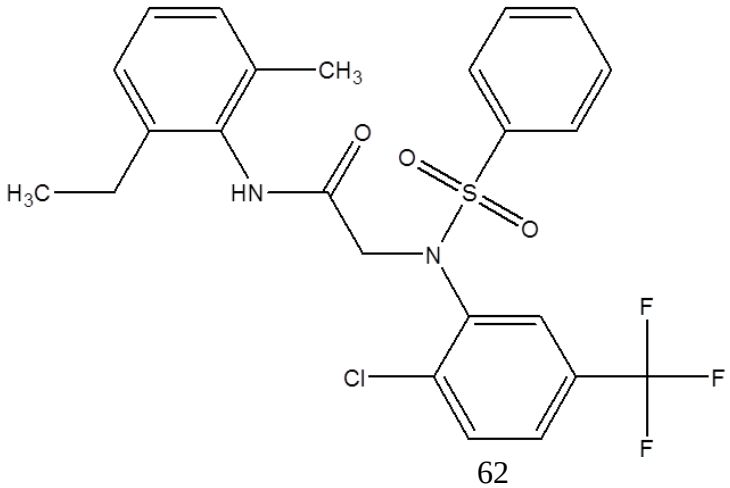


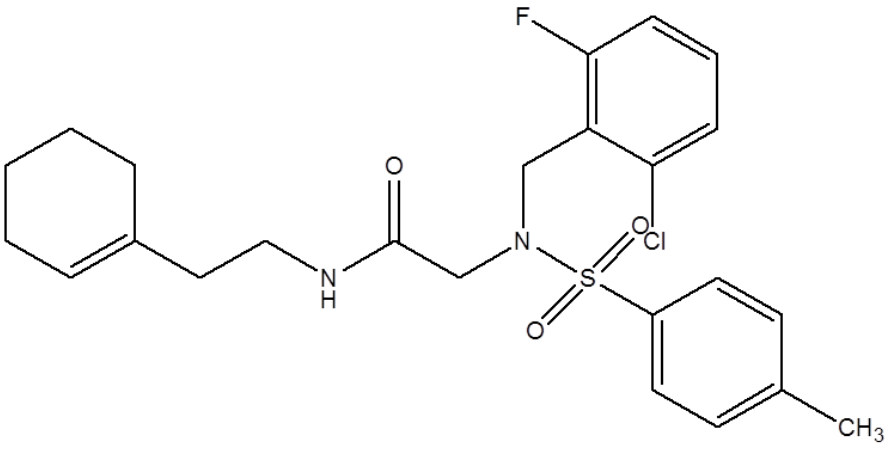


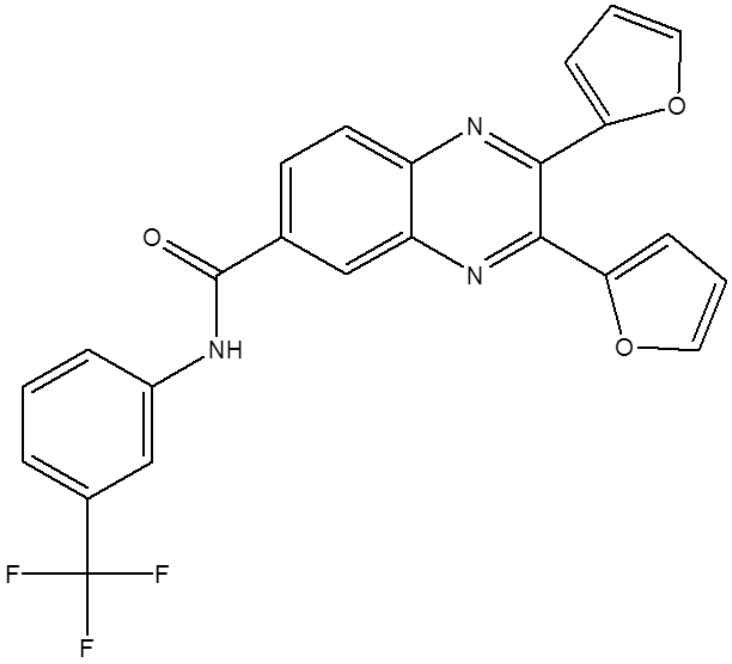


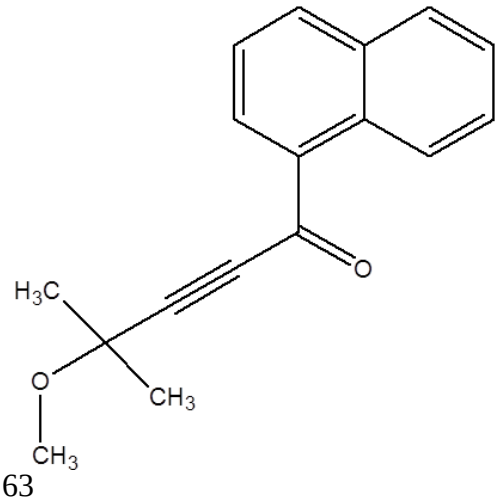


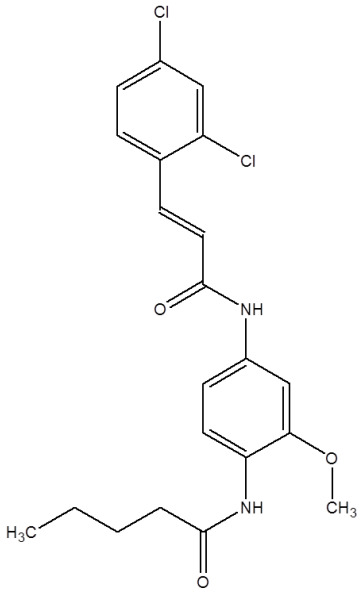





F
O
Cl
N
H
N
S
O
O
CH3


O
N
N
O
O
NH
F
F
F


O
H3C
O
CH3
CH3


Cl
Cl
O
NH
O
NH
CH3
H3C
O